

Title	Genomics approaches to exploit the biotechnological potential of marine sponge-derived <i>Streptomyces</i> spp. isolates
Authors	Almeida, Eduardo L.
Publication date	2019-09
Original Citation	Almeida, E. L. 2019. Genomics approaches to exploit the biotechnological potential of marine sponge-derived <i>Streptomyces</i> spp. isolates. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2019, Eduardo Leao de Almeida. - <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">https://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Download date	2023-05-05 13:34:13
Item downloaded from	<a href="http://hdl.handle.net/10468/9923">http://hdl.handle.net/10468/9923</a>

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



**Genomics approaches to exploit the biotechnological potential  
of marine sponge-derived *Streptomyces* spp. isolates**

Thesis presented by

**Eduardo Leao de Almeida**

for the degree of

**Doctor of Philosophy**

**University College Cork**

**School of Microbiology**

Head of School: Prof Paul O'Toole

Supervisor: Prof Alan D.W. Dobson

2019

# Table of Contents

Table of Contents.....	2
Declaration.....	7
Acknowledgements.....	9
Abstract.....	10
Chapter 1.....	13
1. Marine <i>Streptomyces</i> isolates as reservoirs of compounds of biotechnological and clinical interest.....	14
1.1. Historical perspective of natural products.....	14
1.2. The “Golden Age of Antibiotics” and <i>Streptomyces</i> -derived bio-active compounds.....	17
1.3. The importance of the discovery of new antibiotics.....	20
1.4. Exploring diverse environmental niches in the search for new bio-active molecules.....	21
2. Strategies to identify novel compounds and enzymes from marine <i>Streptomyces</i> isolates.....	26
2.1. The genetics basis for the production of secondary metabolites in bacterial species.....	26
2.2. Genome mining and the state-of-the-art of bioinformatics-assisted compounds discovery.....	30
2.3. The “One Strain Many Compounds” (OSMAC) approach combined with genome mining.....	34
2.4. The discovery and characterisation of the surugamides family of compounds.....	37
3. PETases and the enzymatic hydrolysis of synthetic polyesters.....	41
3.1. Plastic pollution in terrestrial and marine environments.....	41
3.2. PET hydrolases sources, characterisation and mechanisms.....	43
4. References.....	48

Chapter 2.....	61
1. Abstract.....	62
2. Introduction.....	63
3. Material and methods.....	67
3.1. Bacterial strains, maintenance and differential growth assessment....	67
3.2. Genome sequencing, assembly and annotation.....	67
3.3. Comparative genomics.....	68
3.4. Accession numbers.....	70
4. Results and discussion.....	71
4.1. Genome sequencing and assembly.....	71
4.2. Determining the closest terrestrial type-strain relative for the marine sponge-derived isolates.....	75
4.3. Phenotype, morphology, and differential growth assessment.....	79
4.4. Prediction of secondary metabolite biosynthetic gene clusters (smBGCs).....	81
4.5. Comparative genomics.....	87
4.5.1. Analysis of orthologous genes.....	87
4.5.2. Orthology analysis of smBGC-associated genes.....	89
4.5.3. Groups of orthologous genes commonly present in the marine sponge-derived isolates.....	90
4.5.3.1. Resistance to osmotic stress.....	94
4.5.3.2. Antimicrobial compounds production and resistance.....	96
4.5.3.3. ABC transporters.....	99
4.5.3.4. Transcriptional regulation.....	100
4.5.3.5. Genomic evolution through horizontal gene transfer....	101
4.5.3.6. Eukaryotic-like proteins and potential host interaction	103
4.5.4. ENA gene pool genes commonly present in other environmental <i>Streptomyces</i> isolates.....	104



5. Conclusion.....	107
6. References.....	109
7. Supplementary material.....	128
Chapter 3.....	139
1. Abstract.....	140
2. Introduction.....	141
3. Results and discussion.....	145
3.1. Multi-locus sequence analysis and taxonomy assignment of the <i>Streptomyces</i> sp. SM17 isolate.....	145
3.2. Analysis of groups of orthologous genes in the <i>albidoflavus</i> phylogroup.....	148
3.3. Prediction of secondary metabolites biosynthetic gene clusters in the <i>albidoflavus</i> phylogroup.....	151
3.4. Phylogeny and gene synteny analysis of <i>sur</i> BGC homologs.....	154
3.5. Growth, morphology, phenotype and metabolism assessment of <i>Streptomyces</i> sp. SM17 in complex media.....	159
3.6. Differential production of surugamide A by <i>Streptomyces</i> sp. SM17 and <i>S. albidoflavus</i> J1074.....	160
4. Conclusions.....	165
5. Materials and methods.....	167
5.1. Bacterial strains and nucleotide sequences.....	167
5.2. Phylogenetic analyses.....	167
5.3. Prediction of secondary metabolites biosynthetic gene clusters.....	168
5.4. Gene synteny analysis.....	168
5.5. Diagrams and figures.....	169
5.6. Strains culture, maintenance, and secondary metabolites production.....	169
5.7. Metabolic profiling, compound isolation and chemical structure analysis.....	170

6. References.....	172
7. Supplementary material.....	181
Chapter 4.....	183
1. Abstract.....	184
2. Introduction.....	185
3. Material and methods.....	190
3.1. Data sets.....	190
3.2. Bacterial strains.....	192
3.3. Protein homology search and phylogeny analysis.....	192
3.4. PCL plate clearing assay.....	192
3.5. Protein structure analysis, modelling and molecular docking.....	193
3.6. Heterologous expression.....	193
4. Results and Discussion.....	196
4.1. <i>In silico</i> screening of PETase-like proteins in <i>Streptomyces</i> genomes and enzyme activity assessment.....	196
4.2. Protein structure analysis.....	200
4.5.5. Amino acid sequence, conserved residues and domains.....	200
4.5.6. Protein three-dimensional structure prediction and molecular docking.....	203
4.3. Heterologous expression of SM14est.....	205
5. Conclusion.....	208
6. References.....	209
7. Supplementary material.....	218
Chapter 5.....	223
1. General discussion.....	224
1.1. Marine sponge-derived <i>Streptomyces</i> species isolates: the importance of morphological, phenotypical, and genetic characterisation, and their potential as sources of novel bio-active compounds and enzymes.....	224

## Table of Contents

1.2. Genome mining and the OSMAC approach as powerful methods for the identification of natural products in <i>Streptomyces</i> isolates.....	228
1.3. The potential of marine-derived <i>Streptomyces</i> isolates to harbour enzymes with synthetic polyesters-degrading activities.....	231
2. Concluding remarks.....	237
3. References.....	240
Appendix I.....	245

## **Declaration**

I, Eduardo Leao de Almeida, hereby certify that this thesis and the work presented in it are my own, and have not been submitted for any other degree, either at University College Cork or elsewhere.

---

Eduardo Leao de Almeida

*“Two roads diverged in a wood, and I—  
I took the one less traveled by,  
And that has made all the difference.”*

Excerpt from “The Road Not Taken,” by Robert Frost

## Acknowledgements

It has been a long road to reach to this point, although it felt more like a roller coaster at times. The PhD itself took four years and a few months; however, in fact, all the hard work, choices, struggles and achievements in the course of my academic career have led me here; alongside with a number of amazing people that supported me throughout this journey, making everything possible.

First and foremost, I would like to thank my supervisor and mentor, Prof Alan Dobson, for accepting me as his PhD student in this project. His guidance and support were essential for every aspect of the work presented here, and for shaping me as a better researcher. My sincerest thanks for believing in me and for dedicating so much time in helping me, for which I will be forever grateful.

Thanks to my sponsor, the Brazilian National Council for Scientific and Technological Development (CNPq), for the unique opportunity awarded to me, which enabled all this to happen. Thanks also to the excellent staff at University College Cork and the School of Microbiology for all the assistance provided to me during this time.

Furthermore, thanks to all my colleagues and co-workers in our research group: Beatriz Gil-Pulido, Erik Borchert, Felipe Carrillo, Jamie FitzGerald, Lekha Menon Margassery, Lynn Naughton, Maureen Wada Ihua, Sloane Patry, and Stephen Jackson, for all the help, collaborations, and coffee breaks through this period.

Finally, I would like to thank my family and friends, especially my parents and greatest inspirations Maria do Socorro and José Filho; my brothers Rodrigo, Maurício and Henrique; my grandmothers Jacy and Marlene; and all the Leão and the Almeida family; for the unconditional love that kept me going through the good and also the difficult times. To Eduardo and Verena, which will one day be able to read this: this is also to you. And a very special thanks to my girlfriend, Franzi, whose love, patience, encouragement and support were invaluable, especially towards the final push to conclude this thesis.

## Abstract

1           Members of the *Streptomyces* genus are widely known for their capability in  
2   producing compounds of pharmacological, clinical, and biotechnological interest,  
3   being the source of approximately a third of all the antibiotics that have been  
4   identified to date. However, the discovery of natural products with antimicrobial  
5   activities has declined following the so-called “Golden Age of Antibiotics” (1940s-  
6   1950s), particularly due to the common re-discovery of previously known  
7   compounds. Thus, natural products discovery research has shifted towards  
8   investigating diverse environmental niches, such as marine ecosystems, mangroves,  
9   and symbiotic communities of insects and sponges, resulting in the discovery of a  
10   variety of previously unidentified compounds of pharmacological interest; including  
11   those isolated from marine-derived *Streptomyces* species. However, in despite of  
12   their relevance as producers of potentially novel bio-active molecules with  
13   pharmacological, clinical and biotechnological interest, marine-derived  
14   *Streptomyces* isolates are still rather underexplored and under-characterized,  
15   particularly those found in association with marine sponges.

16           In the studies presented in this thesis, various state-of-the-art methodologies  
17   related to genome mining and bioinformatics-based pipelines, together with  
18   molecular and synthetic biology, were employed and proved to be extremely useful  
19   in helping to uncover the biotechnological potential of marine sponge-derived  
20   *Streptomyces* isolates. These studies essentially aimed at a) genetically characterizing  
21   marine sponge-derived *Streptomyces* spp. isolates and their potential to produce  
22   novel secondary metabolites, as shown in Chapter 2; b) to *in silico* identify, isolate,  
23   and quantify a secondary metabolite produced by a marine sponge-derived  
24   *Streptomyces* isolate, together with genetically characterizing its genome-encoded  
25   biosynthetic gene cluster (BGC), as reported in Chapter 3; and c) to perform an *in*  
26   *silico* screening of a novel polyesterase from a marine sponge-derived *Streptomyces*

27 isolate, followed by heterologous protein expression in an *E. coli* host, as  
28 demonstrated in Chapter 4.

29 In Chapter 2, two of the first complete genomes from marine sponge-derived  
30 *Streptomyces* spp. isolates were determined, namely from *Streptomyces* sp. SM17  
31 and *Streptomyces* sp. SM18. The high-quality data provided in this study allowed  
32 for a reliable prediction of secondary metabolites biosynthetic gene clusters (BGCs)  
33 in their genomes, which determined that these isolates possess a variety of BGCs  
34 potentially encoding for the production of known compounds, and also potentially  
35 new molecules. Differential growth assessment determined that the marine isolates  
36 SM17 and SM18 grew and differentiated better in the presence of salts in the culture  
37 medium, when compared to their phylogenetically determined closely-related  
38 terrestrial relatives, namely *S. albidoflavus* J1074 (referred to as *S. albus* J1074 in  
39 Chapter 2) and *S. pratensis* ATCC 33331, respectively. Comparative genomics  
40 allowed for the identification of a proposed environmental niche adaptations (ENA)  
41 gene pool, which included genes related to osmotic stress defence, transcriptional  
42 regulation; symbiotic interactions; antimicrobial compound production and  
43 resistance; ABC transporters; together with horizontal gene transfer and defence-  
44 related features. These results shed new light on some of the genetic traits possessed  
45 by these marine sponge-derived isolates, and on how these might be linked to  
46 secondary metabolites production, and further highlighted their importance for the  
47 discovery of potentially novel natural products.

48 In Chapter 3, the previously unreported capability of the *Streptomyces* sp.  
49 SM17 to produce surugamides has been described. Surugamides are a family of  
50 compounds that have been previously reported to possess antitumor and antifungal  
51 activities. This was performed employing genome mining, which allowed for the  
52 identification of the surugamides BGC (*sur* BGC) in the SM17 genome, and  
53 analytical chemistry techniques for compound isolation and quantification.



54 Phylogenomics analyses provided novel insights with respect to the distribution and  
55 conservation of the *sur* BGC at a genetic level, and provided evidence that the *sur*  
56 BGC might have had a marine origin. Additionally, when comparing the  
57 surugamide A production capabilities of a marine isolate (strain SM17) with a  
58 terrestrial relative (strain J1074) employing a “One Strain Many Compounds”  
59 (OSMAC)-based cultivation approach, the *Streptomyces* sp. SM17 isolate was shown  
60 to produce higher levels of surugamide A in all the conditions tested for. These  
61 findings may provide important insights towards a better characterisation, improved  
62 production and industrial development of this family of compounds.

63 In Chapter 4, the capability of marine sponge-derived *Streptomyces* spp.  
64 isolates to degrade synthetic polyesters was investigated. This was based on the fact  
65 that these microorganisms might have developed mechanisms to assimilate  
66 components of micro-plastics, which are now believed to be ubiquitous in marine  
67 ecosystems and pose as one of the top environmental problems that society faces  
68 today. Using 15 known PET hydrolases (PETases) as references, including the  
69 *Ideonella sakaiensis* 201-F6 PETase, *in silico* screening was performed to determine  
70 the presence of homologs to these reference PETase enzymes in 52 *Streptomyces*  
71 genome sequences (of which 29 were derived from marine ecosystems). The best  
72 candidate identified, namely the SM14est protein from the marine sponge-derived  
73 *Streptomyces* sp. SM14, was *in silico* characterised with respect to its amino acid  
74 sequence and predicted three dimensional structure, and was subsequently  
75 heterologously expressed in an *E. coli* host. This allowed for the confirmation of the  
76 polyesterase activity possessed by the SM14est enzyme, via a polycaprolactone (PCL)  
77 plate-clearing assay. Better characterising, identifying sources, and determining  
78 methods for improved protein expression are essential steps towards the  
79 development of biotechnological applications and industrial processes employing  
80 this family of enzymes, such as new plastic waste processing technologies.

# Chapter 1

## Introduction

## 1. Marine *Streptomyces* isolates as reservoirs of compounds of biotechnological and clinical interest

### 1.1. Historical perspective of natural products

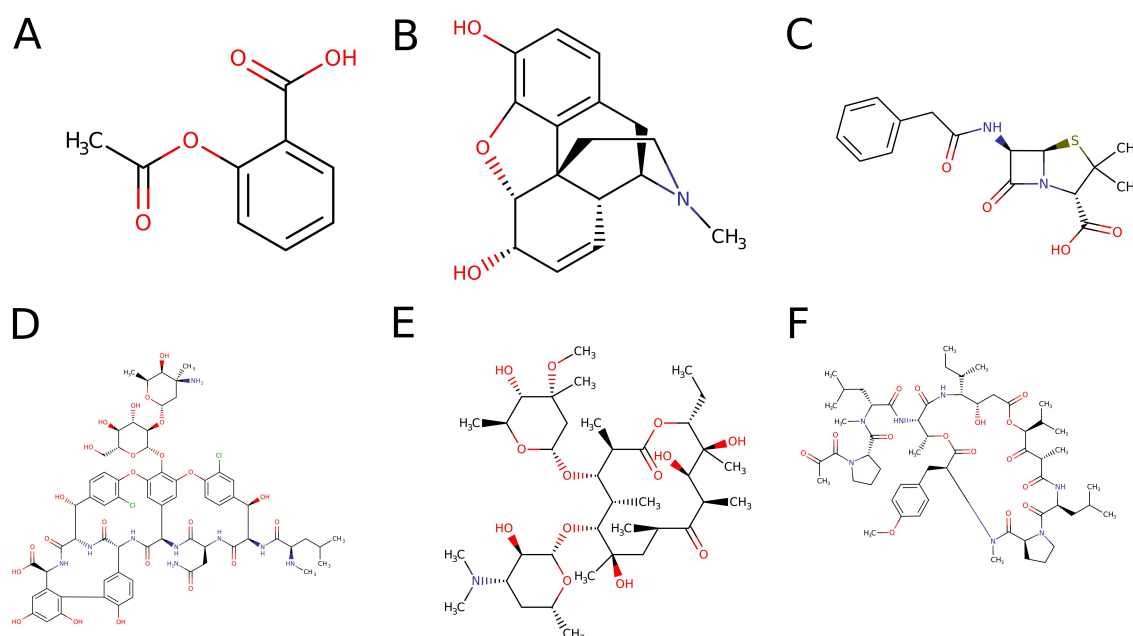
The discovery and applications of natural products (NPs) have been important in the development of pharmacology and medicine, and for the development of human society over many numbers of years (Cragg and Newman, 2013; Dias et al., 2012). NPs are also referred to as “secondary metabolites”, or “specialised metabolites”, and this class of molecule generally encompasses compounds that are usually not required for growth, development, or reproduction of the producing organism; and are thus rather considered to be involved in mechanisms required for adaptation to the environment, communication (e.g. quorum sensing), and/or defence (Bibb, 2005; Cragg and Newman, 2013; Dias et al., 2012; Luckner, 2014). In contrast, “primary metabolites” are those involved in the biosynthesis and breakdown of molecules which are considered essential to all living organisms, such as nucleic acids, proteins, carbohydrates, and lipids (Dias et al., 2012; Hodgson, 2004; Luckner, 2014). Therefore, whereas primary metabolism is found to be present in all varieties of macro- and microorganisms, the ability to produce different types of secondary metabolites is believed to be much more limited, hence certain types of secondary metabolites are only found in certain organisms from particular environmental niches (Bibb, 2005; Cragg and Newman, 2013; Dias et al., 2012; Luckner, 2014).

Plants-derived NPs, in particular, have been historically exploited for numerous applications in medicine, with the earliest records dating from 2,600 B.C., in Mesopotamia; with the use of around 1,000 of these plant-derived compounds having been documented, including oils from plants that are used even to this day, e.g. *Cupressus sempervirens* and *Commiphora* species, in the treatment of

107 inflammation, coughs, and colds (Cragg and Newman, 2013; Dias et al., 2012). The  
108 usage of plant-based compounds was also extended to other ancient civilisations;  
109 which have also been documented, for example, in the Egyptian pharmaceutical  
110 record “Ebers Papyrus” which dates from 1,500 B.C.; the Chinese Materia Media  
111 from 1,100 B.C.; the records of medicinal herbs by the Greek physician Dioscorides  
112 from 100 A.D.; while monasteries in Europe together with the Arabs were  
113 responsible for preserving much of our knowledge about medicinal plants  
114 throughout the 5<sup>th</sup> to 12<sup>th</sup> centuries, in the Dark and Middle Ages (Cragg and  
115 Newman, 2013; Dias et al., 2012). Among important drugs that are derived from  
116 plant NPs, there is aspirin (acetylsalicylic acid) (Figure 1A), which is derived from  
117 the salicin – a secondary metabolite isolated from *Salix alba* L. (willow tree) (Dias  
118 et al., 2012; Newman et al., 2000; Schrör, 2016). Morphine (Figure 1B) is also a classic  
119 example of a plant-derived NPs of commercial and clinical relevance, which was  
120 originally isolated from *Papaver somniferum* L. (opium poppy) (Dias et al., 2012;  
121 Newman et al., 2000).

122         Other than plant-derived NPs, important sources of bio-active compounds  
123 also include fungi, bacteria, and, more recently, marine organisms such as sponges  
124 and algae (Carroll et al., 2019; Cragg and Newman, 2013; Dias et al., 2012; Newman  
125 and Cragg, 2016). With respect to fungal-derived NPs, there is penicillin (Figure 1C),  
126 isolated from *Penicillium notatum*, which has an importance for modern medicine  
127 that cannot be overestimated and which has undoubtedly saved countless lives  
128 (Dias et al., 2012; Fleming, 1929, 1944; Gaynes, 2017; Newman et al., 2000). Among  
129 clinically relevant bacterial-derived NPs, there are the glycopeptide antibiotic  
130 vancomycin (Figure 1D) and the macrolide antibiotic erythromycin (Figure 1E)  
131 (Haight and Finland, 1952; McCormick et al., 1955). The former, which was isolated  
132 from *Amycolatopsis orientalis*, possesses bio-activity against Gram-positive and  
133 Gram-negative bacteria and also fungi; while the latter, which was isolated from

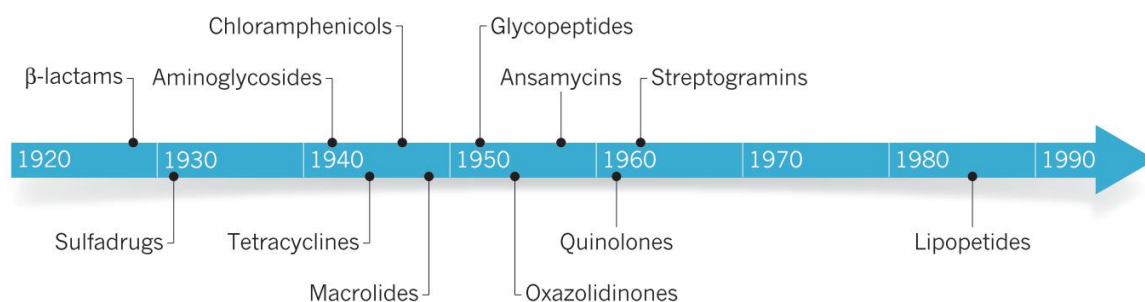
*Saccharopolyspora erythraea*, is commonly used to treat respiratory tract infections, and possesses bio-activity against Gram-positive bacilli and cocci (Dias et al., 2012). More recently, secondary metabolites derived from marine organisms have been reported to be particularly effective as anticancer agents (Carroll et al., 2019; Dias et al., 2012). The secondary metabolite plitidepsin (Figure 1F), for example, was isolated from the tunicate *Aplidium albicans* and has been shown to be effective in the treatment of various types of cancers (Bravo et al., 2005; Carroll et al., 2019; Dias et al., 2012; Mateos et al., 2005).



**Figure 1:** Chemical structure of the natural products A) acetylsalicylic acid (aspirin); B) morphine; C) penicillin G; D) vancomycin; E) erythromycin; F) plitidepsin. Structures obtained from <https://www.drugbank.ca/> (Wishart et al., 2018).

## 1.2. The “Golden Age of Antibiotics” and *Streptomyces*-derived bio-active compounds

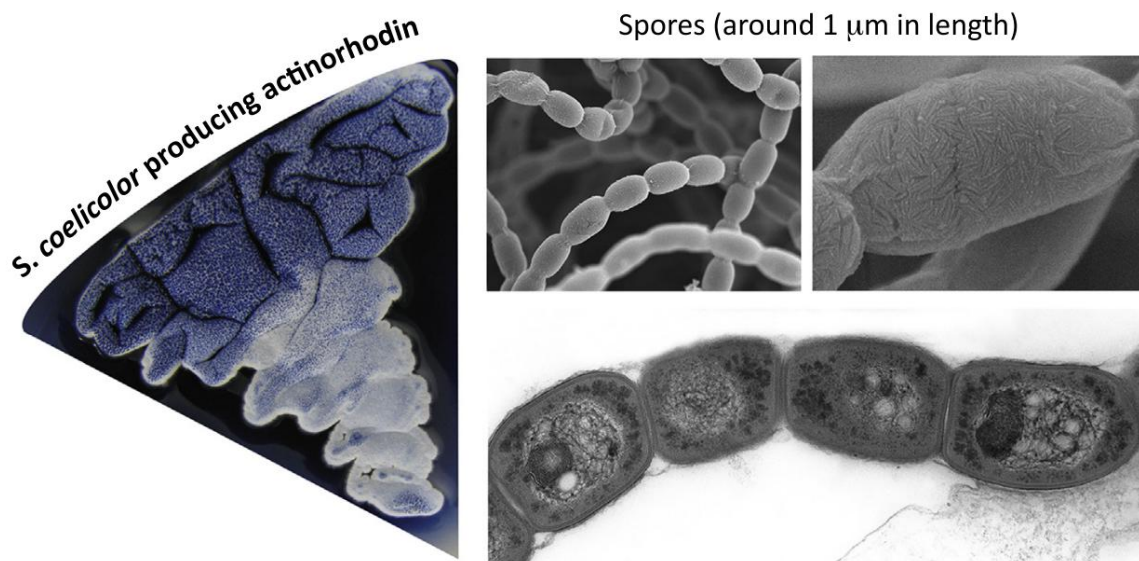
The discovery of perhaps the first and most important NP derived from a microorganism, namely the aforementioned penicillin from the fungus *Penicillium notatum*, which was discovered by Fleming in 1929 (Fleming, 1929, 1944; Gaynes, 2017), initiated a new era in modern medicine, in the so-called “Golden Age of Antibiotics” (1940s-1950s) (Figure 2) (Fernandes, 2006; Lewis, 2012; Mohr, 2016). This prompted the investigation of various microorganisms, in particular bacteria belonging to the Actinomycetales order, as potential sources of what later became extremely valuable bio-active secondary metabolites for clinical applications and for the pharmaceutical industry; such as those with antimicrobial, immunosuppressive, and anticancer activities (Carroll et al., 2019; Challinor and Bode, 2015; Lewis, 2012).



**Figure 2:** Timeline representing the discovery of antibiotics molecules, with a clear decrease in the discovery of new families of compounds after the 1960s (Lewis, 2012).

Among these bacteria which were shown to possess the capacity to produce bio-active compounds of clinical, commercial and pharmacological interest, members of the *Streptomyces* genus – which at that point had been isolated mainly from soil – were determined to be highly prolific producers of antibiotics, being the

166 source of up to two thirds of all known antibiotics to date (Mohr, 2016; Watve et al.,  
167 2001). The *Streptomyces* genus (Actinomycetales order) consists of Gram-positive,  
168 filamentous, aerobic, and mycelial bacteria, with a complex development,  
169 morphological differentiation, and multicellular behaviour (Figure 3) (Hoskisson  
170 and van Wezel, 2019; Law et al., 2018; McCormick and Flärdh, 2012). *Streptomyces*  
171 species possess long linear chromosomes (6-10 Mbp) with high G+C content,  
172 usually above 70%; which is considerably higher than *Bacillus* species, for example,  
173 which commonly possess around a 40% G+C content (Hoskisson and van Wezel,  
174 2019; Lee et al., 2019; Muto and Osawa, 1987). Members of the *Streptomyces* genus  
175 are well-known for their exceptional capabilities of producing secondary  
176 metabolites with various bio-activities, particularly those with antimicrobial  
177 activity (Clardy et al., 2006; Watve et al., 2001). The first couple of antibiotics to be  
178 isolated from *Streptomyces* species were streptothricin, in 1942 (Waksman and  
179 Woodruff, 1942), and streptomycin, in 1944 (Schatz et al., 1944); the latter of which  
180 is considered to be responsible for triggering a more systematic screening approach  
181 for the discovery of novel bio-active compounds derived from this genus (Clardy et  
182 al., 2006; Watve et al., 2001). It has been estimated that around 3,000 antibiotics  
183 molecules have to date been identified that are produced by the Actinomycetales  
184 order, of which 90% are derived from *Streptomyces* species (Clardy et al., 2006;  
185 Watve et al., 2001). Additionally, in a conservative estimate, 150,000 compounds  
186 have been proposed to be produced by *Streptomyces* species; thus, only a very  
187 small fraction – around 1-3% – of the full potential number of compounds that may  
188 be produced by *Streptomyces* have been identified to date (Clardy et al., 2006;  
189 Watve et al., 2001).



**Figure 3:** Morphology of *Streptomyces coelicolor* (Hoskisson and van Wezel, 2019).

Among clinically and commercially relevant compounds derived from *Streptomyces* species, others include the antibiotics chloramphenicol, which was isolated from *Streptomyces venezuelae* (Bartz, 1948; Ehrlich et al., 1948); daptomycin, isolated from *Streptomyces roseosporus* (Raja et al., 2003); fosfomycin, isolated from *Streptomyces fradiae* (Hendlin et al., 1969); lincomycin, isolated from *Streptomyces lincolnensis* (Schaffer et al., 1963); neomycin, isolated from *Streptomyces fradiae* (Waksman and Lechevalier, 1949); the aforementioned streptomycin, isolated from *Streptomyces griseus* (Schatz et al., 1944); and tetracycline, isolated from *Streptomyces rimosus* and *Streptomyces aureofaciens* (Putnam et al., 1953). Other clinically relevant compounds that are also produced by *Streptomyces* species, include the anticancer compounds doxorubicin (Arcamone et al., 1969) and daunorubicin (Dubost et al., 1964), which were isolated from *Streptomyces peucetius*; and also streptozotocin, isolated from *Streptomyces achromogenes* (Vavra et al., 1959).



### 1.3. The importance of the discovery of new antibiotics

Although, as previously mentioned, numerous secondary metabolites were identified and ultimately commercialised in the course of the so-called “Golden Age of Antibiotics”, a steady decrease in the number of novel compounds identified was observed throughout subsequent years, mostly due to the re-discovery of previously known molecules (Fernandes, 2006; Lewis, 2012). This phenomenon led to a reluctance by the pharmaceutical industry, to invest in research aimed at the discovery of new bio-active NPs; and even to the belief that apparently all the antibiotics molecules may have been discovered (Fernandes, 2006). The latter, of course, has been proven not to be the case, particularly since 1) only a very small fraction of the microbial world has to date been explored since it has been estimated that only 1–2% of environmental bacteria are culturable (Nesme et al., 2016; Torsvik et al., 1990; Wade, 2002), an issue which could be addressed with the development of new culture-independent technologies, such as metagenomics (Chen et al., 2019); and 2) it has been estimated that the *Streptomyces* genus alone could potentially be able to produce tens of thousands of compounds that have not yet been discovered (Watve et al., 2001).

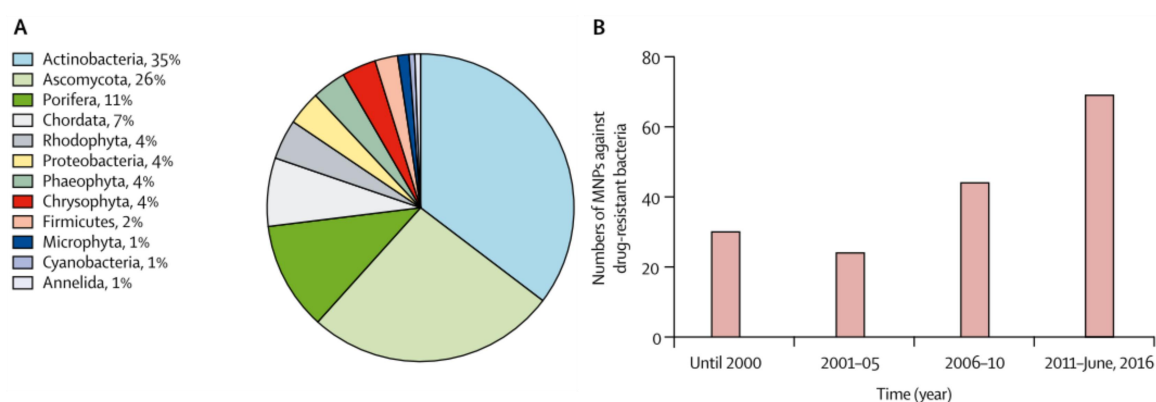
However, new and important challenges have recently arisen particularly in a clinical context, with respect to the treatment of bacterial infections. In the past few decades, an increase in bacterial resistance to broad spectrum antibiotics has been reported (Pendleton et al., 2013; Sirijan and Nitaya, 2016; Tommasi et al., 2015), which has consequently increased the necessity of finding novel antimicrobial compounds, and in particular those that possesses alternative mechanisms of bacterial growth inhibition; as one of the top priorities in contemporary scientific research. In 2009, it has been estimated that multidrug-resistant (MDR) bacterial infections caused 25,000 extra-deaths per year in Europe, while a more recent study estimated that MDR bacteria-related infections could lead to a number as high as

10 million deaths per year by 2050 (Abat et al., 2018; O'Neill, 2014). These MDR bacterial isolates are commonly responsible for life-threatening nosocomial infections, with a particular concern surrounding those belonging to a group of microorganisms which have been described as the ESKAPE pathogens – an acronym that refers to Gram-positive and Gram-negative bacterial species that include *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* spp. (Pendleton et al., 2013; Sirijan and Nitaya, 2016; Tommasi et al., 2015).

#### 1.4. Exploring diverse environmental niches in the search for new bio-active molecules

The aforementioned decline in the discovery of novel bio-active compounds together with the increased need to find novel molecules – particularly those with antimicrobial activity due to the emergence of multi-drug resistant bacteria – has then led to a shift in efforts aimed at the discovery of novel NPs (Pendleton et al., 2013; Tommasi et al., 2015). Concerted efforts started to focus on the investigation of microorganisms derived from diverse environmental niches, which had previously been neglected when compared to soil isolates; such as those isolated from marine ecosystems, mangroves, and in particular those found in association with terrestrial and marine invertebrates, such as insects and marine sponges (Carroll et al., 2019; Chevrette et al., 2019; Dias et al., 2012; Hassan et al., 2017; Kemung et al., 2018; Manivasagan et al., 2014). These led to the successful identification and isolation of completely novel small molecules with antibacterial, antifungal, and anticancer activities; including those with inhibitory activity against multi-drug resistant ESKAPE pathogens (Figure 4) (Andryukov et al., 2019; Liu et al., 2019b; Schinke et al., 2017). From marine-derived *Bacillus* species, for example, the compounds bogorol A and loloatin B have been reported, which possess inhibitory activity against methicillin-resistant *Staphylococcus aureus*

(MRSA) and vancomycin-resistant enterococcal strains (VRE) (Barsby et al., 2001; Gerard et al., 1996). The compounds stachyin B, stachybocin A, stachybocin B, and ilicicolin – all of which were isolated from the marine fungus *Stachybotrys* sp. MF347 – have also been reported to have inhibitory activity against MRSA (Liu et al., 2019b; Wu et al., 2014). Hence, it became clear that microorganisms derived from marine ecosystems might provide a valuable resource in the treatment of multi-drug resistant infections.



**Figure 4:** Graphical representation of 167 marine-derived natural products active against drug-resistant bacteria. A) Phylum-wise distribution. B) Time-wise distribution (adapted from Liu et al., 2019).

Among marine-derived sources of novel bio-active small molecules, marine sponges have been reported to host several groups of microorganisms such as fungi and bacteria which are capable of producing quite unique secondary metabolites with a range of bio-activities, including some with antimicrobial and anticancer properties (Carroll et al., 2019). To date, sponge bacteria-derived antimicrobial compounds have been identified from 35 different genera, with *Streptomyces* being the most predominant genus; being the source of approximately 30% of the compounds identified to date (Indraningrat et al., 2016). Examples of compounds which have recently been isolated from marine sponge-

279 associated *Streptomyces* isolates (Table 1) that have been reported to inhibit some  
280 of the ESKAPE pathogens include mayamycin, which has activity against  
281 *Staphylococcus aureus*, and streptophenazines G and K, which have activity  
282 against *Staphylococcus epidermis* and *Bacillus subtilis* (Table 1). Both of these  
283 compounds are produced by *Streptomyces* sp. HB202 isolated from the sponge  
284 *Halichondria panicea* (Table 1) (Schneemann et al., 2010); together with the  
285 compound 2-pyrrolidone, which has activity against *Staphylococcus aureus*,  
286 *Escherichia coli*, and *Klebsiella pneumoniae*, and was found to be produced by  
287 *Streptomyces* sp. MAPS15 isolated from *Spongia officinalis* (Table 1)  
288 (Sathiyarayanan et al., 2014). Additionally, bio-activity screening followed by  
289 genetic characterisation studies performed with 13 *Streptomyces* spp. isolates from  
290 both shallow water and deep-sea sponges have reported antimicrobial activities  
291 against several clinically relevant bacterial and yeast species, including MRSA  
292 (Jackson et al., 2018; Kennedy et al., 2009). These isolates included *Streptomyces* sp.  
293 SM17, which possesses inhibitory activity against *Escherichia coli* NCIMB 12210,  
294 MRSA, and *Candida* species; and *Streptomyces* sp. SM18, which possesses  
295 inhibitory activity against MRSA and *Bacillus subtilis* 1A40; bio-activities which  
296 were determined using deferred antagonism assays (Jackson et al., 2018; Kennedy  
297 et al., 2009). Although further investigation is required in order to isolate specific  
298 bio-active compounds and ultimately commercialise them as novel pharmaceutical  
299 drugs, it is clear that marine sponge-derived *Streptomyces* isolates possess the  
300 potential to produce novel bio-active secondary metabolites with clinical and  
301 biotechnological relevance; and these aforementioned recent studies provide clear  
302 evidence that investigating bacteria such as *Streptomyces* species from these  
303 marine-based environmental niches is a promising field of research.

Sponge	Organism	Compound	Target
<i>Spongia officinalis</i>	<i>Streptomyces</i> sp. MAPS15	2-pyrrolidone	<i>Staphylococcus aureus</i> PC6; <i>Escherichia coli</i> PC1; <i>Klebsiella pneumoniae</i> PC7
<i>Halichondria panicea</i>	<i>Streptomyces</i> sp. HB202	Mayamycin	MRSA; <i>Staphylococcus epidermidis</i> ; <i>Brevibacterium epidermidis</i> ; <i>Dermabacter hominis</i> ; <i>Propionibacterium acnes</i> ; <i>Xanthomonas campestris</i>
		Streptophenazine G	<i>Staphylococcus epidermidis</i> ; <i>Bacillus subtilis</i>
		Streptophenazine K	<i>Staphylococcus epidermidis</i>
<i>Haliclona simulans</i>	<i>Streptomyces</i> sp. SM8	Mixture Kitamycin A or B and Antimycin A3 or A7	<i>Bacillus subtilis</i>
		Antimycin A2, A8, A11 or A17	
		Antimycin A3 or A7	
		Antimycin A2, A8, A11 or A17, antimycin A3 or A7	
<i>Dysidea tupha</i>	<i>Streptomyces</i> sp. RV15	Naphthacene glycoside SF2446A2	<i>Chlamydia trachomatis</i>

304 **Table 1:** Compounds with antimicrobial activity isolated from marine  
305 sponge-derived *Streptomyces* spp. (adapted from Indraningrat et al.,  
306 2016).

307 It has been proposed that environmental niche adaptations possessed by  
308 these marine sponge-derived *Streptomyces* species may play a role, at least in part,  
309 in their ability to produce a diverse range of secondary metabolites (Almeida et al.,  
310 2019; Bibb, 2005), as the abiotic conditions faced by these microorganisms such as  
311 differences in pressure, salinity, pH, light, temperature and oligotrophic conditions

312 are likely to be linked to the biosynthesis of NPs (Abdelmohsen et al., 2014; van der  
313 Meij et al., 2017). In addition to these abiotic stresses, symbiotic relationships may  
314 also be particularly relevant for the production of secondary metabolites in these  
315 marine sponge-derived *Streptomyces* isolates (Chevrette et al., 2019; Kwan et al.,  
316 2014). It has been reported, for example, that the production of antimicrobial  
317 compounds was promoted when the sponge-associated *Streptomyces* ACT-52A  
318 was exposed to the sponge *Aplysilla rosea* (Mehbub et al., 2016). However, the  
319 mechanisms that regulate the production of secondary metabolites in marine  
320 sponge-associated *Streptomyces* species is likely to be quite complex and to date  
321 remains for the most part unresolved. Thus, it is of crucial importance to better  
322 characterise these microorganisms morphologically, phenotypically and genetically.

## 323 2. Strategies to identify novel compounds and enzymes from marine 324 *Streptomyces* isolates

### 325 2.1. The genetics basis for the production of secondary metabolites in bacterial 326 species

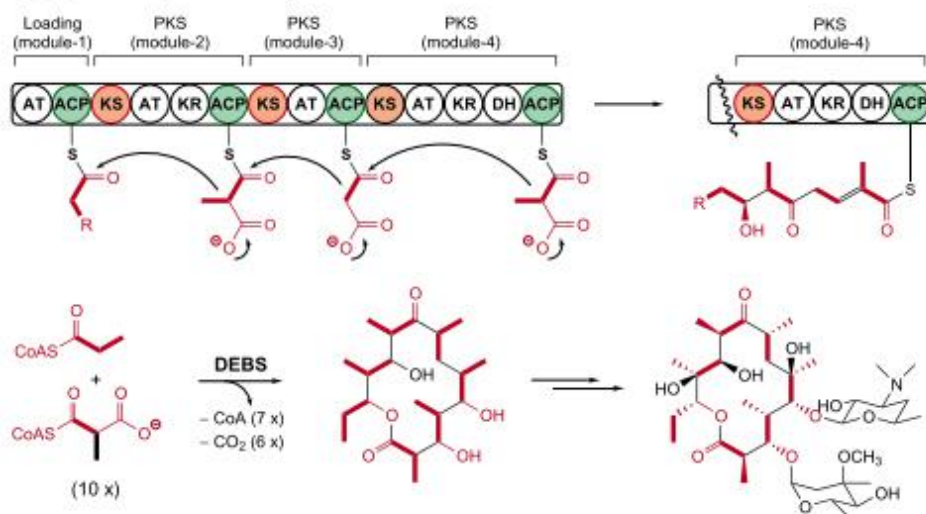
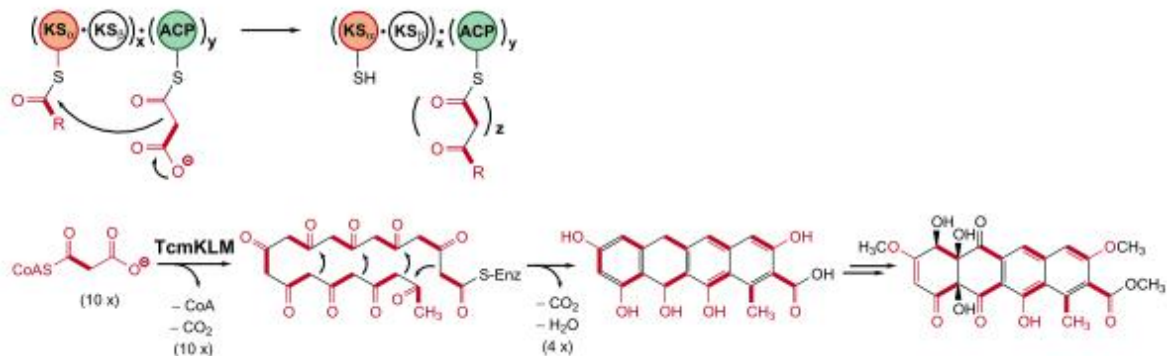
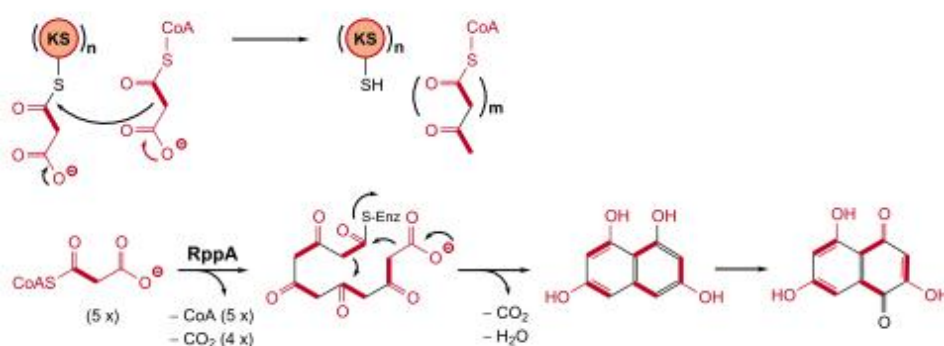
327 The biosynthesis of secondary metabolites is a complex process that, at a  
328 genetic level, is performed by a physical clustering of two or more genes that  
329 together encode for the enzymes involved in the biosynthetic pathways required for  
330 the production of these secondary metabolites (Chen et al., 2019; Medema et al.,  
331 2015). These group of genes are commonly referred to as secondary metabolite  
332 biosynthetic gene clusters (BGCs). In addition to genes encoding enzymes that are  
333 required for the production and assembly of secondary metabolites, BGCs also  
334 contain regulatory, resistance, and transport-related genes (Chen et al., 2019).

335 The main families of compounds produced by BGCs consist of polyketides,  
336 nonribosomal peptides, terpenoids, ribosomally synthesised and post-translationally  
337 modified peptides (RiPPs), saccharides, and alkaloids (Medema et al., 2015). Of these,  
338 the polyketides and nonribosomal peptides are of major research interest as they  
339 commonly comprise of natural products with antibacterial, antifungal, and  
340 immunosuppressant activities (Chen et al., 2019). Polyketides are synthesised by  
341 polyketide synthases (PKSs), while nonribosomal peptide synthases (NRPSs) are  
342 responsible for synthesising nonribosomal peptides, and these BGCs often possess  
343 complex structure and distinct mechanisms of biosynthesis.

344 To date, three main types of bacterial PKSs have been described (Figure 5)  
345 (Robbins et al., 2016; Shen, 2003). Type I PKSs are modular, non-iterative,  
346 multifunctional enzymes, for which each module is responsible for a different  
347 elongation step of the polyketide chain (Figure 5a) (Shen, 2003). The type I PKS 6-  
348 deoxyerythronolide B synthase (DEBS), for example, is involved in the synthesis of

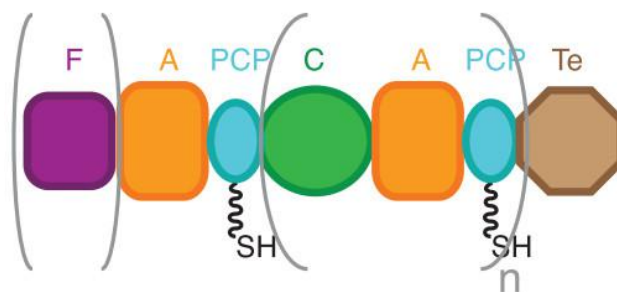
349 the macrolide antibiotic erythromycin (Khosla et al., 2007; Staunton and Wilkinson,  
350 1997; Weber et al., 1991). Type II PKSs, on the other hand, are complexes of multiple  
351 small, discrete enzymes with iterative activities and particular functions (Figure 5b)  
352 (Shen, 2003; Yu et al., 2012). The anthracycline antibiotic tetracenomycin C is an  
353 example of a compound that is produced by a type II PKS (Hutchinson, 1997).  
354 Finally, type III PKSs comprise of self-contained homodimeric enzymes that perform  
355 iterative reactions to assemble polyketide products (Figure 5c) (Shen, 2003; Yu et al.,  
356 2012). In bacterial species, it has been reported that type III PKSs produce  
357 compounds of biological relevance and also provide precursors for the biosynthesis  
358 of other secondary metabolites (Katsuyama and Ohnishi, 2012; Shimizu et al., 2017).  
359 For example, the type III PKS SrsA from *S. griseus* is involved in the biosynthesis  
360 alkylquinone, which has been reported to confer penicillin resistance to this  
361 microorganism (Funabashi et al., 2008; Katsuyama and Ohnishi, 2012). Additionally,  
362 the type III PKS PhlD from *Pseudomonas* is involved in the biosynthesis of 2,4-  
363 diacetylphloroglucinol, which has been reported to possess biocontrol activity  
364 against soil borne fungal plant pathogens (Bangera and Thomashow, 1999;  
365 Katsuyama and Ohnishi, 2012). Besides the clear differences which exist between the  
366 different types of PKSs, all of them possess a ketoacyl synthase (KS) in the form of a  
367 domain(s), for type I PKSs, or as a subunit, for type II and III PKSs (Figure 5) (Shen,  
368 2003).



**(a) Type I PKS (noniterative)****(b) Type II PKS (iterative)****(c) Type III PKS (ACP-independent & iterative)**

**Figure 5:** Examples of (a) type I, (b) type II, and (c) type III PKS structures and mechanisms of biosynthesis. AT: acyl transferase; ACP: acyl carrier protein; KS: ketoacyl synthase; KR: ketoreductase; DH: dehydratase (adapted from Shen, 2003).

373 In contrast, NRPSs are a family of mega-enzymes that contain multiple  
374 domains (adenylation, peptidyl carrier protein, condensation, and thioesterase  
375 domains) that form modules (initiation, elongation and termination modules);  
376 functioning together to incorporate amino acids into the peptide chain (Figure 6)  
377 (Miller and Gulick, 2016; Reimer et al., 2018). Adenylation (A) domains are  
378 responsible for adding the substrate to the nascent peptide chain and usually  
379 consist of approximately 500 amino acid residues; condensation (C) domains  
380 catalyse the bond formation between two substrates and usually consist of  
381 approximately 450 residues; peptidyl carrier protein (PCP) domains are the  
382 smallest NRPS domains and usually consist of 70–90 amino acid residues, and are  
383 responsible for the transfer of substrate and peptide intermediates between  
384 different domains; and finally the thioesterase (Te) domains, which usually consist  
385 of 230–270 amino acid residues, are responsible for releasing the peptide and  
386 freeing the NRPS enzyme (Miller and Gulick, 2016). Initiation modules commonly  
387 consist of A and PCP domains; while elongation modules encompass C domains  
388 together with A and PCP domains; and termination modules possess the Te  
389 domain. Hence, A-PCP-(C-A-PCP)<sub>n</sub>-Te is considered to be the canonical  
390 organisation of a basic NRPS (Figure 6) (Reimer et al., 2018). Examples of NRPS-  
391 encoded secondary metabolites of clinical and biotechnological relevance include  
392 the antibiotics gramicidin A, daptomycin, vancomycin and bacitracin A; the  
393 immunosuppressant cyclosporin A; and the compound surfactin, which can be  
394 employed for the remediation of oil-contaminated soils (Felnagle et al., 2008;  
395 Martínez-Núñez and López, 2016; Reimer et al., 2018).



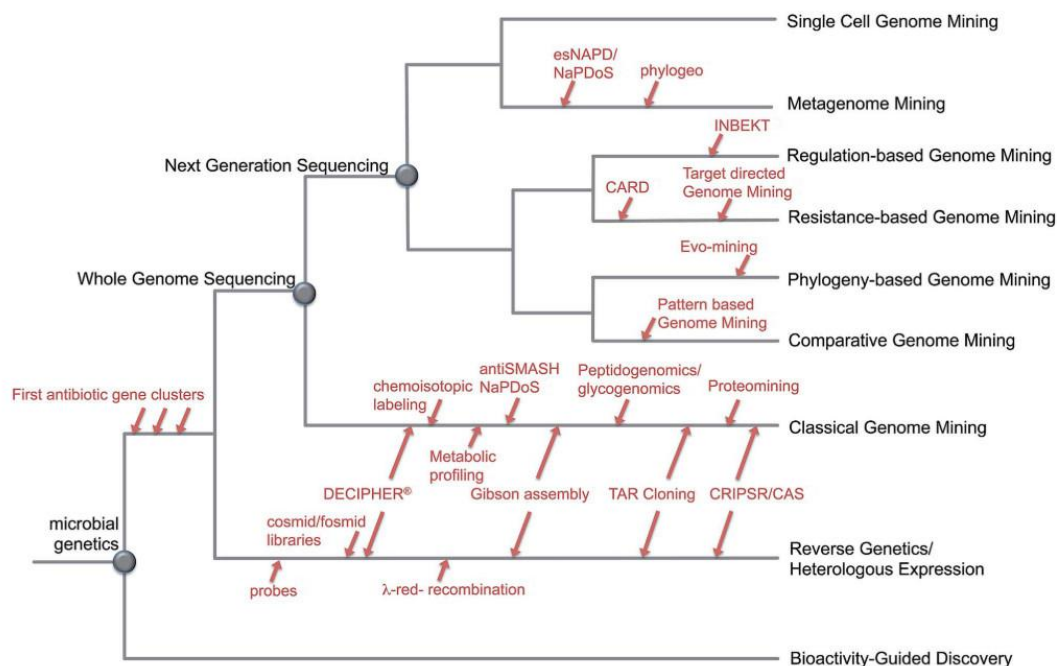
396 **Figure 6:** Canonical domain and module organisation of a generic  
 397 NRPS enzyme. Although represented in the figure, the tailoring  
 398 domain that formylates (F) the N-terminal amino acid is not present  
 399 in all NRPSs (adapted from Reimer et al., 2018).

## 400 2.2. Genome mining and the state-of-the-art of bioinformatics-assisted 401 compounds discovery

402 Traditionally, in the pre-omics era, the discovery of natural products were  
 403 typically performed based on culture-dependent techniques and bio-activity  
 404 screening of plants, fungi and bacteria, coupled with chemistry-based methods  
 405 such as high-performance liquid chromatography (HPLC); mass spectrometry, and  
 406 nuclear magnetic resonance (NMR) for the isolation and characterisation of  
 407 compounds (Chen et al., 2019; Lee et al., 2019; Ziemert et al., 2016). For a number of  
 408 reasons, the use of these approaches alone proved to be relatively limited after  
 409 some time, which resulted at least in part to a decline in the number of novel  
 410 compounds identified following the so-called “Golden Age of Antibiotics”  
 411 (Fernandes, 2006; Lewis, 2012). Technical aspects such as costs and the time  
 412 required for the isolation of compounds; the common rediscovery of previously  
 413 known compounds; together with the fact that culturing environmental  
 414 microorganisms in the laboratory can be very challenging, since most bacteria are  
 415 considered to be “unculturable”; resulted in helping to limit the discovery of novel

416 bio-active compounds when employing strictly *in vitro*-based approaches (Chen et  
417 al., 2019).

418         However, with advances in DNA sequencing, genomics, bioinformatics,  
419 analytical chemistry and synthetic biology, together with the first complete genome  
420 sequence of a *Streptomyces* isolate, namely *Streptomyces coelicolor*, in 2002  
421 (Bentley et al., 2002), a new revolution in the discovery of novel bio-active  
422 compounds began, employing the concept of “genome mining” (Ziemert et al.,  
423 2016). Essentially, genome mining consists on the prediction and isolation of  
424 natural products based on genetic information, even without a chemical structure  
425 being available (Ziemert et al., 2016). As more genetic information regarding  
426 secondary metabolite biosynthetic pathways and their regulation began to be  
427 elucidated, it became increasingly apparent that even previously well-studied  
428 *Streptomyces* isolates could potentially produce different and quite novel  
429 compounds; other than those that had previously been identified and characterised  
430 using chemical-based methods (Ziemert et al., 2016). Since then, our knowledge  
431 with respect to the genetic basis for the production of secondary metabolites – with  
432 the discovery and characterisation of BGCs – has vastly improved (Ziemert et al.,  
433 2016). This has led to the development of a number of *in silico* tools and  
434 bioinformatics-based pipelines, together with synthetic biology and high-  
435 throughput biochemical methods, which has helped to optimise the discovery of  
436 potentially novel and unique compounds; many of which would have been  
437 virtually impossible to uncover if not for these new approaches (Figure 7) (Chen et  
438 al., 2019; Lee et al., 2019; Ziemert et al., 2016).



**Figure 7:** The evolution of genome mining throughout the omics era (adapted from Ziemert et al., 2016).

Classical “genome mining”, which is based on the search for genes encoding enzymes such as PKSs or NRPSs, which are known to be involved in the biosynthesis of secondary metabolites, is among the most broadly used genomics-based approach currently employed in small molecule/natural products discovery. In its basic form, classical genome mining utilises sequence- or profile-based comparison tools, such as BLAST (Altschul et al., 1990; Camacho et al., 2009) and HMMer (Zhang and Wood, 2003), respectively; while more advanced, specialised, and automated tools have also been developed, including the antiSMASH, BAGEL, and the ClusterFinder programs (Blin et al., 2019; Cimerancic et al., 2014; de Jong et al., 2006). These more sophisticated tools are also employed for comparative genome mining, which is a method that employs the analysis and comparison of complete genomes or genomic regions, in a search for partial or complete BGCs, instead of single genes and enzymes. Among these tools, antiSMASH is one of the most prominently employed programs for the prediction of BGCs in the genomes

of bacterial isolates (Blin et al., 2017, 2019). This program allows for the comparison of genome sequences with other known BGC-encoding sequences present in other microorganisms and in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) curated database (Epstein et al., 2018; Medema et al., 2015). This allows for the identification of known and also potentially new BGCs, and is particularly useful for *Streptomyces* species. In fact, around a third of all the curated BGCs available in the MIBiG database to date belong to *Streptomyces* species, which further highlights the importance of the *Streptomyces* genus in the discovery of secondary metabolites. Comparative genome mining methods can be further strengthened when coupled with other *in silico* pipelines such as sequence similarity network analysis, for example by using programs such as the BiG-SCAPE (Navarro-Muñoz et al., 2018), which clusters predicted BGCs by antiSMASH into gene cluster families according to sequence similarity. Such an approach has proven to be very useful in assessing the novelty of potential BGCs and in prioritising isolates and compounds for further investigation (Blin et al., 2017; Ziemert et al., 2016).

Additional interesting and innovative modern methods of genome mining include phylogeny-based mining, which can be performed using tools such as the EvoMining program (Sélem-Mojica et al., 2018); and resistance gene-based mining, which can be performed using the Antibiotic Resistant Target Seeker (ARTS) tool (Alanjary et al., 2017). Additionally, updated information regarding the latest bioinformatics tools for the discovery of secondary metabolites can be found in The Secondary Metabolite Bioinformatics Portal, at <http://www.secondarymetabolites.org> (Weber and Kim, 2016).

The aforementioned bioinformatics-based approaches have proven to be extremely useful in the discovery and activation of the so-called “silent” or “cryptic” BGCs, i.e. those predicted BGCs that are apparently not being expressed under standard laboratory culture conditions. One such example is *Streptomyces*

482 *albidoflavus* J1074, where the use of chemical stress elicitors, i.e. the cytotoxins  
483 etoposide and ivermectin, promoted the production of compounds previously  
484 unknown to be produced by this strain, namely surugamides molecules and their  
485 derivatives, which were determined to be encoded by a BGC which until that point  
486 was considered to be transcriptionally silent in this isolate (Xu et al., 2017).

487 Omics-based approaches have also allowed the investigation of the potential  
488 for the production of secondary metabolites by uncultured organisms present in  
489 different microbiomes, through genome and metagenome sequencing (Blin et al.,  
490 2017; Chen et al., 2019; Cragg and Newman, 2013; Lee et al., 2019; Ziemert et al.,  
491 2016). For example, in a recent study, functional screening of a human microbiome  
492 metagenomic library has led to the identification of a BGC encoding the bacteriocin  
493 colicin V with inhibitory activity against *E. coli*, which may be involved in the  
494 ecology of human-associated bacteria, and could also possess a therapeutic role  
495 (Cohen et al., 2018).

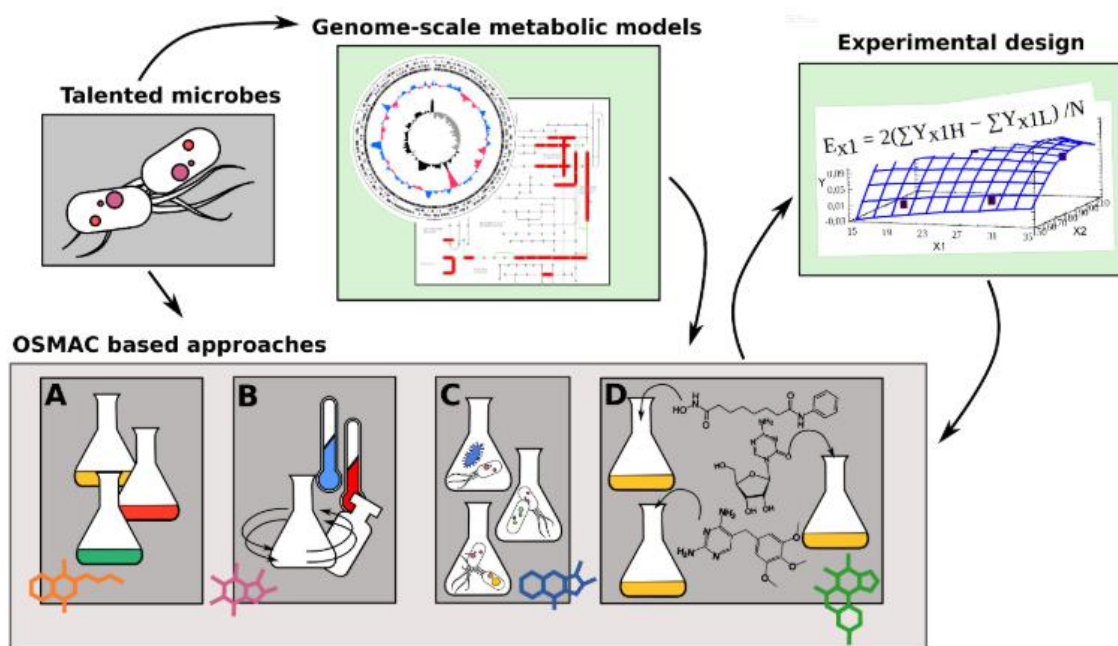
### 496 **2.3. The “One Strain Many Compounds” (OSMAC) approach combined with** 497 **genome mining**

498 The “One Strain Many Compounds” (OSMAC) approach, which was  
499 proposed and conceptualised by Zeeck and collaborators, is based on more  
500 classical approaches employing chemical, physical and culture-dependent methods  
501 in the discovery of novel secondary metabolites (Bode et al., 2002). Essentially, this  
502 method is based on observations that the production of secondary metabolites can  
503 be promoted through exposing the producing strains to stressful culture conditions;  
504 thus systematic changes in simple cultivation parameters, such as the media  
505 composition, culture aeration, and the type of culture vessel, may promote the  
506 production of an increased number of compounds from a single microorganism;  
507 with this approach being particularly useful for *Streptomyces* species (Bibb, 2005;

508 Bode et al., 2002; Chiang et al., 2011; Goodfellow and Fiedler, 2010; Pan et al., 2019;  
509 Romano et al., 2018). The OSMAC approach is a method that is relatively simple to  
510 implement but which can result very useful outcomes, particularly with respect to  
511 the activation of the previously mentioned “silent” BGCs, and hence promoting the  
512 production of compounds which were formerly not produced in some microbial  
513 isolates – even those well-described strains (Chiang et al., 2011; Pan et al., 2019;  
514 Romano et al., 2018; Xu et al., 2017).

515       Typical parameters that can be systematically manipulated in OSMAC-  
516 based experiments include: 1) changes in the nutrients available in the culture  
517 medium, such as changes in carbon, nitrogen, sulphur, and phosphorus sources,  
518 and trace elements; 2) changes in physical parameters, such as temperature, salinity,  
519 pH, aeration and shaking conditions, and culture vessel types; 3) applying  
520 chemical elicitors to the culture medium; and 4) the use of co-cultivation regimes,  
521 in an attempt to simulate biotic interactions found in nature, such as co-culturing  
522 microbial isolates with pathogenic bacteria, or with eukaryotic hosts, in the case of  
523 symbiotic isolates (Figure 8) (Bode et al., 2002; Chiang et al., 2011; Pan et al., 2019;  
524 Romano et al., 2018).





**Figure 8:** Basic implementation of the OSMAC principle, combined with omics approaches. A, B, C, and D represent the employment of different nutrient regimes, physical parameters, co-cultivation, and chemical elicitors, respectively (adapted from Romano et al., 2018).

The OSMAC approach employing different culture media has proven to be particularly useful in investigating novel secondary metabolites produced by marine-derived *Streptomyces* species (Goodfellow and Fiedler, 2010). For example, a significant variation in the metabolic profile of *Streptomyces* sp. YB104, which was isolated from deep-sea sediment, was observed when the isolate was grown in different complex media, which led to the identification of a novel bio-active compound, namely inthomycin B, which belongs to a family of compounds with antimicrobial, herbicidal and anticancer activities (Wu et al., 2018). Also, considerable variations in the levels of the antibiotic SBR-22 being produced by the marine sediment isolate *Streptomyces psammoticus* BT-408, which possesses inhibitory activity against MRSA, has been reported when different carbon sources were present in the culture media (Sujatha et al., 2005). Additionally, levels of

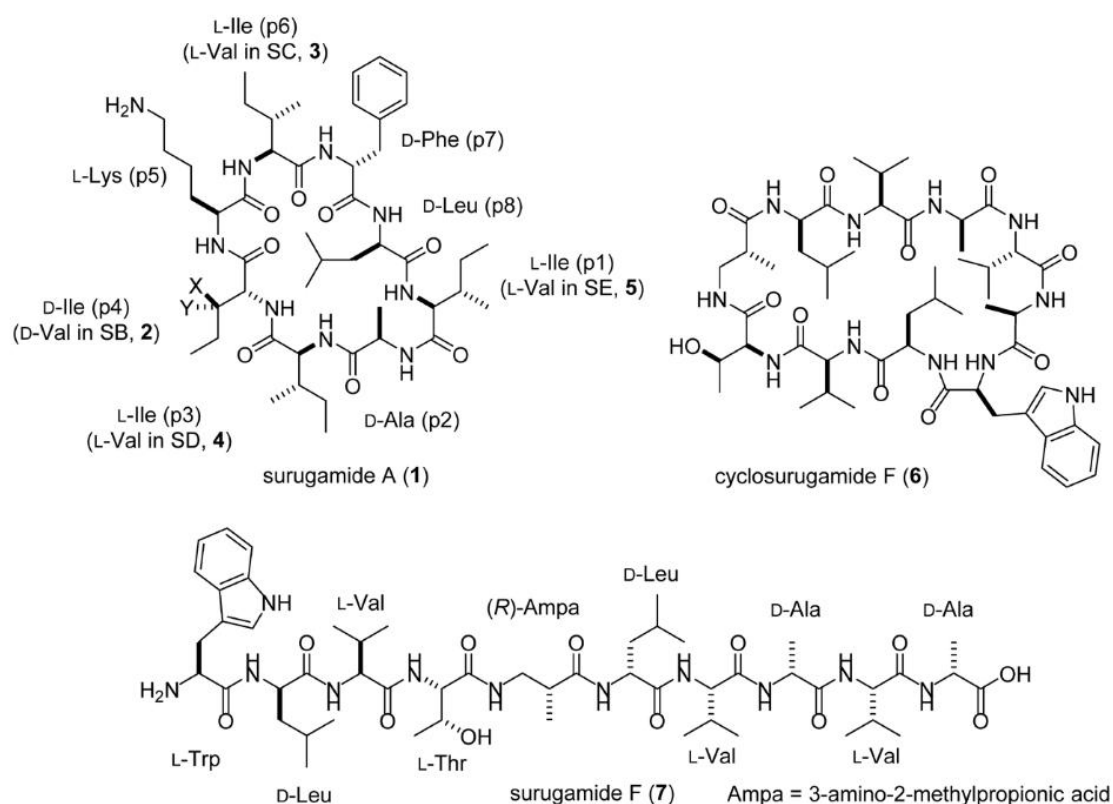
oxygen in the culture have been reported to influence the secondary metabolites production profile in the marine-derived *Streptomyces* sp. CNQ-525, for which hypoxic conditions shifted the production of the antibiotic napyradiomycin towards its intermediate 8-aminoflaviolin, a compound which may potentially function as an endogenous extracellular electron shuttle (Gallagher et al., 2017).

Ultimately, the OSMAC principle can further strengthen genome mining-based natural products discovery, and vice-versa, since it may provide novel insights in metabolic pathways and regulatory mechanisms which can be linked to the genetic information which is available regarding the producing strain. At the same time, prior information on the BGC present in the producing strain – provided by genome mining approaches – can serve as a “blueprint” describing the types of compounds that may potentially be produced by the subject microorganism, and hence provide insights into which culture conditions that may induce production of certain compounds.

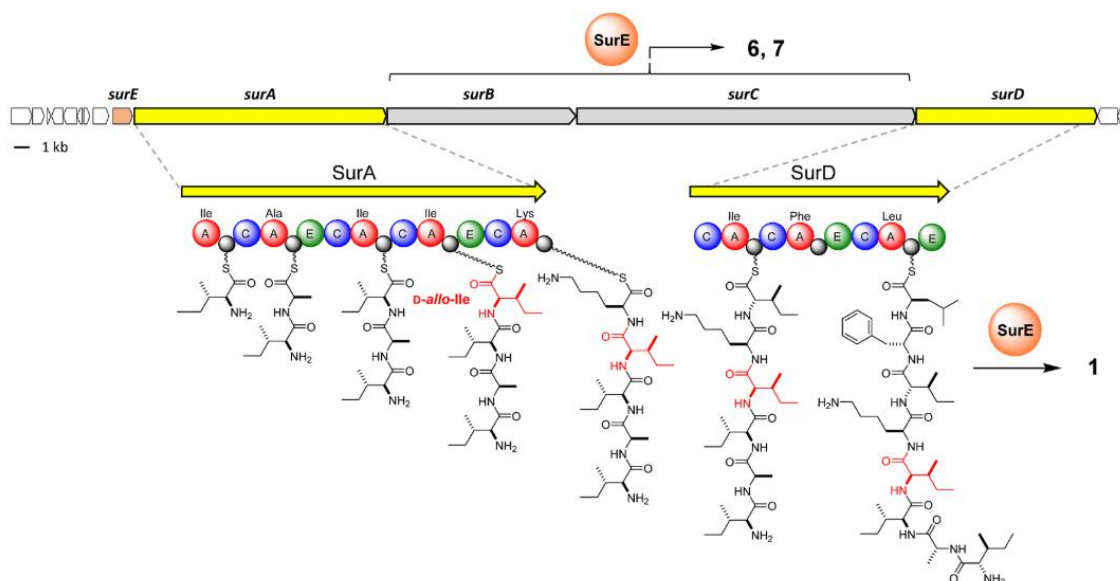
## **2.4. The discovery and characterisation of the surugamides family of compounds**

The recently discovered surugamides family of compounds is a very good example of novel secondary metabolites of clinical relevance, whose discovery involved the application of chemistry-based and culture-dependent methods, together with genomics-based and state-of-the-art synthetic biology approaches (Matsuda et al., 2019b; Ninomiya et al., 2016; Takada et al., 2013; Xu et al., 2017). Surugamides A–E are cyclic octapeptides which have been reported to possess cathepsin B inhibitory activity – an anticancer target – and that were originally identified, isolated and characterised from the marine-derived *Streptomyces* sp. JAMM992 via culture-dependent methods, coupled with liquid chromatography–mass spectrometry (LC-MS) analyses (Figure 9) (Takada et al., 2013). Subsequent research (Ninomiya et al., 2016), which involved genome sequencing and genome

567 mining for BGCs using antiSMASH (Blin et al., 2019), together with genetic  
 568 manipulation generating knockout strains, revealed not only the main biosynthetic  
 569 genes involved in the biosynthesis of surugamides (namely the NRPS genes  
 570 *surABCD*) (Figure 10), but also allowed for the identification of the new compound  
 571 surugamide F, a linear decapeptide. In fact, this study determined an  
 572 unprecedented organisation of a BGC, in which the genes *surAD* were involved in  
 573 the production of surugamides A–E, while the *surBC* genes were involved in the  
 574 production of surugamide F – findings which were only possible due to the  
 575 combination of *in vitro* and *in silico* methodologies.



576 **Figure 9:** Surugamides chemical structures, where 2–5 indicate the  
 577 surugamides B–E, respectively. X = H, Y = Me, p = position (Matsuda et  
 578 al., 2019b).



579 **Figure 10:** The *sur* BGC and the biosynthesis of surugamides, where 1  
 580 indicates the products surugamides A–E, encoded by the *surAD* genes; 6 the  
 581 cyclosurugamide F, and 7 the surugamide F, both which are encoded by the  
 582 *surBC* genes (Matsuda et al., 2019b).

583 Interestingly, similar BGCs to the previously described surugamides BGC  
 584 (*sur* BGC) were also identified in other *Streptomyces* isolates, including the well-  
 585 described *Streptomyces albidoflavus* J1074 (Xu et al., 2017), however, this BGC was  
 586 considered to be “silent” in this strain. Genomics analyses, coupled with high-  
 587 throughput screening employing chemical stress elicitors, allowed for the  
 588 induction of production of surugamides by the isolate, and also for the  
 589 identification of completely novel compounds also produced by the *sur* BGC, such  
 590 as the surugamide I, which also possesses cathepsin B inhibitory activity, and acyl-  
 591 surugamide A, which possesses antifungal activity. This study also employed  
 592 molecular biology and synthetic biology methodologies that allowed for a better  
 593 characterisation of metabolic pathways and genes involved in the production of  
 594 surugamides, identifying the transcriptional repressor *surR* which was reported to  
 595 play an important role in the regulation of production of surugamides.

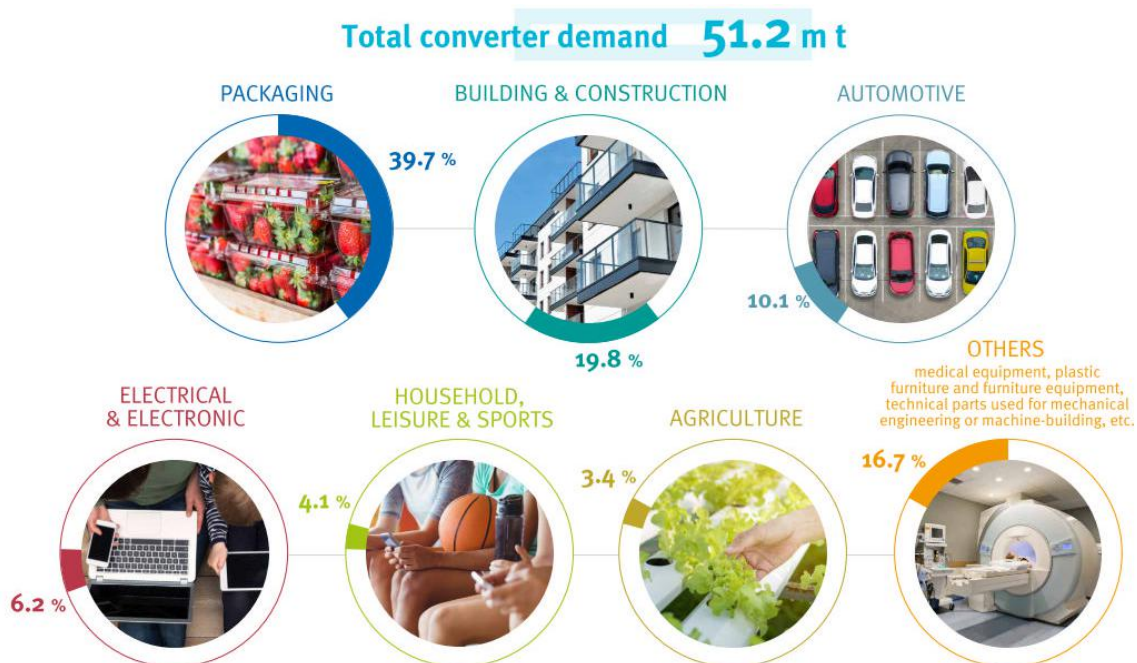
596           More recent studies which further investigated the biosynthesis of  
597 surugamides also applied chemistry-based approaches coupled with genomics and  
598 synthetic biology in order to better characterise genes, enzymes, and metabolic  
599 pathways involved in the production of these molecules. These studies revealed,  
600 for example, another gene which is also involved in the tailoring of these  
601 compounds, namely *surE*, which encodes a thioesterase involved in the cyclisation  
602 of the cyclic surugamides molecules (Kuranaga et al., 2018; Matsuda et al., 2019b,  
603 2019a; Thankachan et al., 2019; Zhou et al., 2019).

604           Taken together, these studies highlight the power of genomics-based  
605 approaches coupled with culture-dependent, chemistry and synthetic biology  
606 methods, employed to identify completely novel molecules of clinical interest, such  
607 as the surugamides, even when applied to well characterised isolates, such as  
608 *Streptomyces albidoflavus* J1074.

### 3. PETases and the enzymatic hydrolysis of synthetic polyesters

#### 3.1. Plastic pollution in terrestrial and marine environments

Synthetic polyesters, or plastics, such as polyethylene terephthalate (PET), are widely used in our society due to their advantageous physico-chemical properties such as their bio-inertia, together with their low production cost. These characteristics are particularly useful for the packaging industry, a sector that contributes to around 40% of the plastics demand in Europe (Figure 11) (Lebreton and Andrady, 2019; PlasticsEurope, 2018). However, as a consequence of these advantageous properties, many synthetic polyesters are highly recalcitrant to biodegradation, and in turn became an enormous problem not only to the environment but to our society as a whole (Wei and Zimmermann, 2017a).



**Figure 11:** Plastics converter demand in Europe in 2017 (adapted from PlasticsEurope, 2018).

622           It has been estimated that, in 2017 alone, worldwide plastics production was  
623 around 348 million tonnes; and that at the current rate, it is estimated that plastics  
624 production could double in the next 20 years (Lebreton and Andrady, 2019;  
625 PlasticsEurope, 2018). Although increased efforts in improving plastic waste  
626 management have helped to decrease the problem of plastic pollution, there is still  
627 a considerable amount of mismanaged plastic waste that ends up in the  
628 environment. In 2015 alone, it has been estimated that 60-90 million metric tonnes  
629 of mismanaged plastic waste were produced globally, and that this amount could  
630 triple by 2060 (Lebreton and Andrady, 2019).

631           Plastic pollution becomes particularly concerning when it reaches the  
632 marine environment, for which it has been estimated that, in 2010 alone, 4.8-12.7  
633 metric tonnes of plastic waste was disposed of in the oceans. Given that these levels  
634 only considered coastal countries, this value could in fact be much larger (Jambeck  
635 et al., 2015). Following recent research, additional issues must be taken into account  
636 when considering plastic pollution in the oceans. It has been reported that, for  
637 example, plastic debris undergoes fragmentation into micro-plastics (< 5mm), due  
638 to weathering effects such as sunlight, and these are believed to be ubiquitous in  
639 many soil and aquatic environments (Geyer et al., 2017; Lebreton and Andrady,  
640 2019). These micro-plastics have been reported to be commonly ingested by  
641 animals, which are believed to transfer pollutants along the food chain, including  
642 to seafood species consumed by humans (Santillo et al., 2017).

643           Mechanical recycling, which involves collection, sorting, washing and  
644 grinding, is currently the most commonly applied method for plastic waste  
645 recycling (Ragaert et al., 2017). A huge challenge for mechanical recycling is,  
646 however, the presence of impurities in the post-consumer plastic waste (Drzyzga  
647 and Prieto, 2019). Alternatively, chemical recycling has also been applied to  
648 improve plastic waste management processes, with the advantage that the plastic

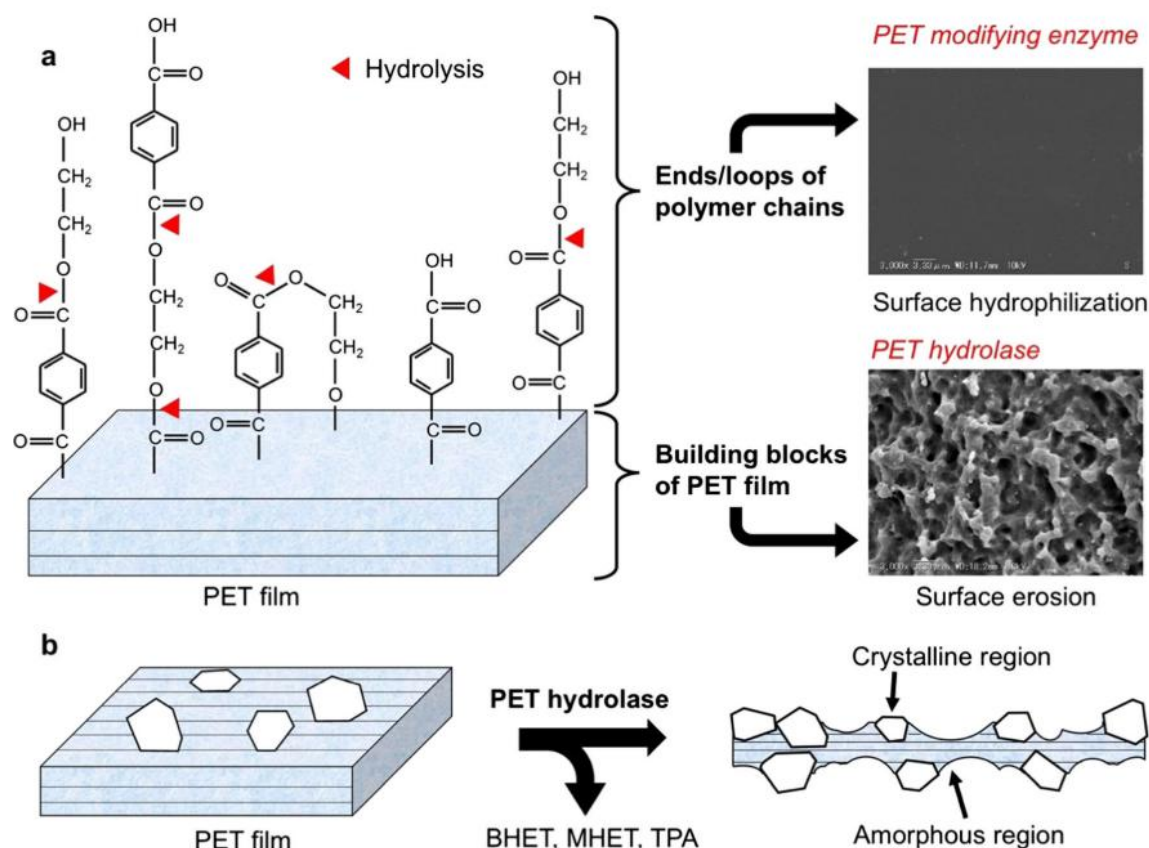
649 polymers can also be converted into raw materials for the synthesis of chemicals,  
650 fuels, and virgin plastics (Drzyzga and Prieto, 2019). However, these chemical  
651 processes can be very costly and usually require the use of high temperatures and  
652 toxic chemicals (Awaja and Pavel, 2005; Wei and Zimmermann, 2017b).

### 653 3.2. PET hydrolases sources, characterisation and mechanisms

654 More recently, a potential alternative for a more efficient and environment-  
655 friendly approach for the recycling of post-consumer plastic waste has been  
656 proposed, namely the enzymatic hydrolysis of plastics (Drzyzga and Prieto, 2019;  
657 Wei and Zimmermann, 2017a, 2017b). A few bacterial-derived enzymes with the  
658 capacity to degrade synthetic polyesters, including PET, have been identified in the  
659 past decade (Kawai et al., 2019; Wei and Zimmermann, 2017a). These were mainly  
660 identified in thermophilic actinomycetes, particularly in the genus *Thermobifida*,  
661 and the enzymes involved commonly comprise of members of the cutinases,  
662 lipases, and esterases families (Silva et al., 2011; Wei and Zimmermann, 2017a). A  
663 ground-breaking discovery in the field occurred when, in 2016, Yoshida and co-  
664 workers isolated a bacterium, namely *Ideonella sakaiensis* 201-F6, which was  
665 capable of degrading and assimilating PET as its major energy and carbon source,  
666 that was isolated from a bottle recycling plant in Sakai, Japan (Yoshida et al., 2016).  
667 The enzyme responsible for the hydrolysis of PET, namely ISF6\_4831, was  
668 classified as a PETase (or PET hydrolase, EC 3.1.1.101), and it has been reported to  
669 possess higher enzymatic activity and substrate specificity for PET than any other  
670 previously described PET hydrolase, being able to degrade PET even at moderate  
671 temperatures of around 30°C (Joo et al., 2018; Yoshida et al., 2016). These findings  
672 prompted research groups worldwide to further characterise this class of enzymes,  
673 together with the metabolic and biochemical processes and pathways involved in  
674 the degradation of PET and other synthetic polyesters (Chen et al., 2018; Han et al.,  
675 2017; Joo et al., 2018; Liu et al., 2018). It has been proposed that PETases act by



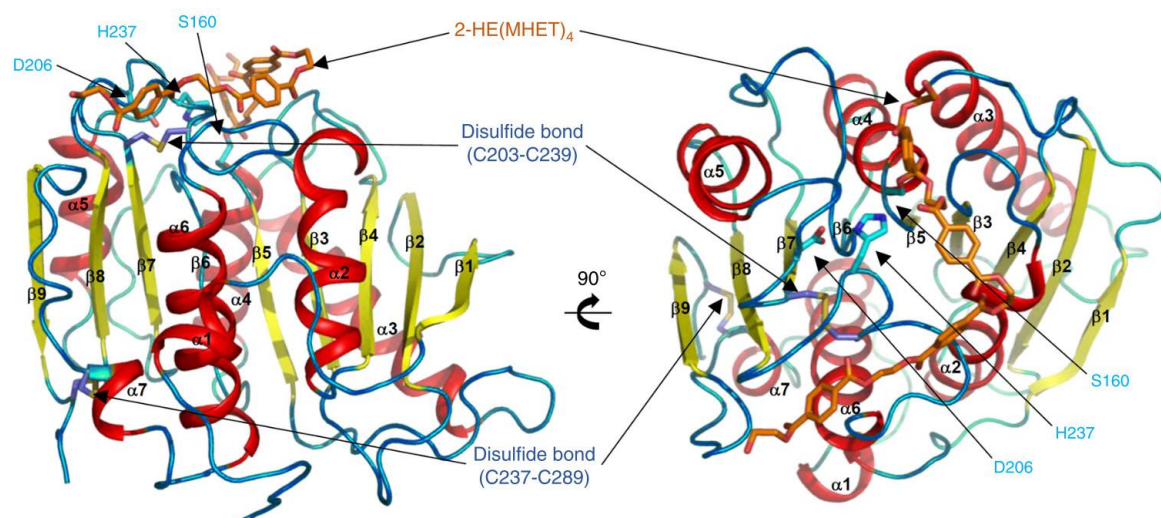
676 catalysing the hydrolytic cleavage of PET, resulting in the production of smaller  
 677 monomeric molecules of ethylene glycol (EG), terephthalate (TPA), mono(2-  
 678 hydroxyethyl) terephthalate (MHET), and bis(2-hydroxyethyl) terephthalate (BHET)  
 679 (Figure 12) (Chen et al., 2018; Han et al., 2017; Kawai et al., 2019).



680 **Figure 12:** General mechanisms of PET hydrolysis by PETase (Kawai et  
 681 al., 2019).

682 The crystal structure of the *Ideonella sakaiensis* 201-F6 PETase ISF6\_4831  
 683 (IsPETase) has determined (Figure 13) (Austin et al., 2018; Han et al., 2017; Joo et al.,  
 684 2018; Liu et al., 2018, 2019a). These studies were aimed at elucidating and gaining  
 685 further insights into the structural characteristics that conferred the efficiency and  
 686 affinity that was observed with respect to the degradation of PET by the IsPETase.  
 687 The enzyme was determined to belong to the  $\alpha/\beta$  hydrolase superfamily (Hotelier

et al., 2004; Nardini and Dijkstra, 1999; Ollis et al., 1992), with a conserved serine hydrolase motif Gly-x1-Ser-x2-Gly (Gly158-Trp159-Ser160-Met161-Gly162), and to be arranged in 9  $\beta$ -sheets and 7  $\alpha$ -helices (Figure 13). By analysing the three-dimensional structure of this enzyme together with employing molecular docking experiments and site-directed mutagenesis assays, some features that could explain the superior hydrolysis of PET by the IsPETase have been proposed (Joo et al., 2018). Firstly, it has been proposed that the IsPETase possesses two sub-sites that facilitate a superior binding to the PET moieties when compared to others PET hydrolases from other microorganisms. Sub-site I (consisting of Tyr87, Met161, Trp185, Ile208), and which seems to be conserved in other PET hydrolases, has been suggested to bind to the first PET MHET moiety, and to be linked to stabilisation of the ligand. On the other hand, the sub-site II (consisting of Thr88, Ala89, Trp159, Ile232, Asn233, Ser236, Ser238, Asn241, Asn244, Ser245, Asn246, Arg280) has been proposed to be involved in accommodating the three MHET moieties of PET, and seems to be a more uncommon structural feature possessed by the IsPETase, therefore being partially responsible for the superior enzymatic activity. Secondly, the IsPETase possesses an extended loop connecting  $\beta$ 8 and  $\alpha$ 6, which has been proposed to provide a conformation that allows improved accommodation of the substrate. Finally, an extra disulfide bond (between Cys203 and Cys239) in the proximity of the catalytic triad (Ser160, Asp206, His237) is present in the IsPETase (Figure 13), which could also be related to a superior catalytic activity of the enzyme when compared to other PETase enzymes (Joo et al., 2018).



**Figure 13:** Three-dimensional structure of the IsPETase, highlighting the two disulfide bonds; the catalytic triad (Ser160, Asp206, His237); and a simulated 2-HE(MHET)<sub>4</sub> molecule mimicking the structure of PET as the binding substrate at the active site (adapted from Joo et al., 2018).

Although also consisting of members of the Actinomycetales order, very little is currently known about the potential of *Streptomyces* species to degrade synthetic polyesters, unlike the previously mentioned *Thermobifida* genus. Some research efforts have previously been employed to investigate the ability of some soil-derived isolates to degrade synthetic polyesters; for example, *Streptomyces* sp. strain MG (Calabia and Tokiwa, 2004; Tokiwa and Calabia, 2004) and *Streptomyces thermoviolaceus* (Chua et al., 2013) have been reported to being able to degrade polycaprolactone (PCL); and *Streptomyces bangladeshensis* 77T-4 which has been shown to degrade poly(D-3-hydroxybutyrate) (PHB) (Hsu et al., 2012). Given the fact that marine-derived *Streptomyces* species, which are well-adapted to diverse environments and nutrient sources – particularly those found in association with marine sponges – are likely to be exposed to micro-plastics which are ubiquitous in the oceans, and hence may have developed mechanisms to degrade and assimilate

729 these synthetic polymer molecules; therefore, it is perhaps reasonable to presume  
730 that enzymes with synthetic polyesters-degrading activity may be present in these  
731 isolates. Thus, further investigation of marine sponge-derived *Streptomyces*  
732 isolates, employing genome mining together with culture-dependent screening  
733 assays and synthetic biology approaches, may prove useful in providing new  
734 insights into marine-derived PETase enzymes and how widespread these enzymes  
735 may be in nature. These PETase enzymes could be extremely useful in the  
736 development of alternative and more environment-friendly processes of plastic  
737 waste recycling and/or bioremediation, thus, helping to alleviate at least partially  
738 the alarming issue of plastic pollution in terrestrial and aquatic environments.

## 4. References

- Abat, C., Fournier, P. -E., Jimeno, M. -T., Rolain, J. -M., and Raoult, D. (2018). Extremely and pandrug-resistant bacteria extra-deaths: myth or reality?. *European Journal of Clinical Microbiology & Infectious Diseases*. 37(9), 1687-1697. doi:10.1007/s10096-018-3300-0.
- Abdelmohsen, U. R., Bayer, K., and Hentschel, U. (2014). Diversity, abundance and natural products of marine sponge-associated actinomycetes. *Nat. Prod. Rep.* 31, 381–399. doi:10.1039/C3NP70111E.
- Alanjary, M., Kronmiller, B., Adamek, M., Blin, K., Weber, T., Huson, D., et al. (2017). The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* 45, W42–W48. doi:10.1093/nar/gkx360.
- Almeida, E. L. de, Carillo Rincon, F., Jackson, S. A., and Dobson, A. D. (2019). Comparative genomics of marine sponge-derived *Streptomyces* spp. isolates SM17 and SM18 with their closest terrestrial relatives provides novel insights into environmental niche adaptations and secondary metabolite biosynthesis potential. *Front. Microbiol.* 10, 1713. doi:10.3389/FMICB.2019.01713.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.
- Andryukov, B., Mikhailov, V., and Besednova, N. (2019). The Biotechnological Potential of Secondary Metabolites from Marine Bacteria. *J. Mar. Sci. Eng.* 7, 176. doi:10.3390/jmse7060176.
- Arcamone, F., Cassinelli, G., Fantini, G., Grein, A., Orezzi, P., Pol, C., et al. (1969). Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from *S. Peucetius* var. *caesius*. *Biotechnol. Bioeng.* 11, 1101–1110. doi:10.1002/bit.260110607.
- Austin, H. P., Allen, M. D., Donohoe, B. S., Rorrer, N. A., Kearns, F. L., Silveira, R. L., et al. (2018). Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc. Natl. Acad. Sci.* 115, E4350–E4357. doi:10.1073/pnas.1718804115.
- Awaja, F., and Pavel, D. (2005). Recycling of PET. *Eur. Polym. J.* 41, 1453–1477. doi:10.1016/J.EURPOLYMJ.2005.02.005.
- Bangera, M. G., and Thomashow, L. S. (1999). Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-

- 774 diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. *J. Bacteriol.* 181,  
775 3155–63.
- 776 Barsby, T., Kelly, M. T., Gagné, S. M., and Andersen, R. J. (2001). Bogorol A  
777 produced in culture by a marine *Bacillus* sp. reveals a novel template for  
778 cationic peptide antibiotics. *Org. Lett.* 3, 437–40.
- 779 Bartz, Q. R. (1948). Isolation and characterization of chloromycetin. *J. Biol. Chem.*  
780 172, 445–50.
- 781 Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R.,  
782 James, K. D., et al. (2002). Complete genome sequence of the model  
783 actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147.  
784 doi:10.1038/417141a.
- 785 Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Curr.*  
786 *Opin. Microbiol.* 8, 208–215. doi:10.1016/j.mib.2005.02.016.
- 787 Blin, K., Kim, H. U., Medema, M. H., and Weber, T. (2017). Recent development of  
788 antiSMASH and other computational approaches to mine secondary  
789 metabolite biosynthetic gene clusters. *Brief. Bioinform.*, 1–11.  
790 doi:10.1093/bib/bbx146.
- 791 Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019).  
792 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.  
793 *Nucleic Acids Res.* doi:10.1093/nar/gkz310.
- 794 Bode, H. B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big Effects from Small  
795 Changes: Possible Ways to Explore Nature's Chemical Diversity.  
796 *ChemBioChem* 3, 619. doi:10.1002/1439-7633(20020703)3:7<619::AID-  
797 CBIC619>3.0.CO;2-9.
- 798 Bravo, S. B., García-Rendueles, M. E. R., Seoane, R., Dosil, V., Cameselle-Teijeiro, J.,  
799 López-Lázaro, L., et al. (2005). Plitidepsin Has a Cytostatic Effect in Human  
800 Undifferentiated (Anaplastic) Thyroid Carcinoma. *Clin. Cancer Res.* 11, 7664–  
801 7673. doi:10.1158/1078-0432.CCR-05-0455.
- 802 Calabia, B. P., and Tokiwa, Y. (2004). Microbial degradation of poly(d-3-  
803 hydroxybutyrate) by a new thermophilic *Streptomyces* isolate. *Biotechnol. Lett.*  
804 26, 15–19. doi:10.1023/B:BILE.0000009453.81444.51.
- 805 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al.  
806 (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.  
807 doi:10.1186/1471-2105-10-421.

- 808 Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2019).  
 809 Marine natural products. *Nat. Prod. Rep.* 36, 122–173.  
 810 doi:10.1039/C8NP00092A.
- 811 Challinor, V. L., and Bode, H. B. (2015). Bioactive natural products from novel  
 812 microbial sources. *Ann. N. Y. Acad. Sci.* 1354, 82–97. doi:10.1111/nyas.12954.
- 813 Chen, C.-C., Han, X., Ko, T.-P., Liu, W., and Guo, R.-T. (2018). Structural studies  
 814 reveal the molecular mechanism of PETase. *FEBS J.* doi:10.1111/febs.14612.
- 815 Chen, R., Wong, H., and Burns, B. (2019). New Approaches to Detect Biosynthetic  
 816 Gene Clusters in the Environment. *Medicines* 6, 32.  
 817 doi:10.3390/medicines6010032.
- 818 Chevrette, M. G., Carlson, C. M., Ortega, H. E., Thomas, C., Ananiev, G. E., Barns, K.  
 819 J., et al. (2019Chevrette). The antimicrobial potential of *Streptomyces* from  
 820 insect microbiomes. *Nat. Commun.* 10, 516. doi:10.1038/s41467-019-08438-0.
- 821 Chiang, Y.-M., Chang, S.-L., Oakley, B. R., and Wang, C. C. C. (2011). Recent  
 822 advances in awakening silent biosynthetic gene clusters and linking orphan  
 823 clusters to natural products in microorganisms. *Curr. Opin. Chem. Biol.* 15,  
 824 137–43. doi:10.1016/j.cbpa.2010.10.011.
- 825 Chua, T.-K., Tseng, M., and Yang, M.-K. (2013). Degradation of Poly( $\epsilon$ -caprolactone)  
 826 by thermophilic *Streptomyces thermoviolaceus* subsp. *thermoviolaceus* 76T-2.  
 827 *AMB Express* 3, 8. doi:10.1186/2191-0855-3-8.
- 828 Cimermanic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C.,  
 829 Mavrommatis, K., et al. (2014). Insights into secondary metabolism from a  
 830 global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421.  
 831 doi:10.1016/j.cell.2014.06.034.
- 832 Clardy, J., Fischbach, M. A., and Walsh, C. T. (2006). New antibiotics from bacterial  
 833 natural products. *Nat. Biotechnol.* 24, 1541–1550. doi:10.1038/nbt1266.
- 834 Cohen, L. J., Han, S., Huang, Y. H., and Brady, S. F. (2018). Identification of the  
 835 Colicin v Bacteriocin Gene Cluster by Functional Screening of a Human  
 836 Microbiome Metagenomic Library. *ACS Infect. Dis.* 4, 27–32.  
 837 doi:10.1021/acsinfecdis.7b00081.
- 838 Cragg, G. M., and Newman, D. J. (2013). Natural products: A continuing source of  
 839 novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* 1830, 3670–3695.  
 840 doi:10.1016/j.bbagen.2013.02.008.
- 841 de Jong, A., van Hijum, S. A. F. T., Bijlsma, J. J. E., Kok, J., and Kuipers, O. P. (2006).  
 842 BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res.* 34,  
 843 W273–W279. doi:10.1093/nar/gkl237.

- 844 Dias, D. A., Urban, S., and Roessner, U. (2012). A Historical Overview of Natural  
845 Products in Drug Discovery. *Metabolites* 2, 303–336.  
846 doi:10.3390/metabo2020303.
- 847 Drzyzga, O., and Prieto, A. (2019). Plastic waste management, a matter for the  
848 “community”. *Microb. Biotechnol.* 12, 66–68. doi:10.1111/1751-7915.13328.
- 849 Dubost, M., Ganter, P., Maral, R., Ninet, L., Pinnert, S., Preudhomme, J., et al. (1964).  
850 Rubidomycin: A New Antibiotic With Cytostatic Properties. *Cancer*  
851 *Chemother. reports* 41, 35–6.
- 852 Ehrlich, J., Gottlieb, D., Burkholder, P. R., Anderson, L. E., and Pridham, T. G. (1948).  
853 *Streptomyces venezuelae*, n. sp., the source of chloromycetin. *J. Bacteriol.* 56,  
854 467–77.
- 855 Epstein, S. C., Charkoudian, L. K., and Medema, M. H. (2018). A standardized  
856 workflow for submitting data to the Minimum Information about a  
857 Biosynthetic Gene cluster (MIBiG) repository: prospects for research-based  
858 educational experiences. *Stand. Genomic Sci.* 13, 16. doi:10.1186/s40793-018-  
859 0318-y.
- 860 Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon,  
861 M. D., et al. (2008). Nonribosomal peptide synthetases involved in the  
862 production of medically relevant natural products. *Mol. Pharm.* 5, 191–211.  
863 doi:10.1021/mp700137g.
- 864 Fernandes, P. (2006). Antibacterial discovery and development—the failure of  
865 success? *Nat. Biotechnol.* 24, 1497–1503. doi:10.1038/nbt1206-1497.
- 866 Fleming, A. (1929). On the Antibacterial Action of Cultures of a *Penicillium*, with  
867 Special Reference to their Use in the Isolation of *B. influenzae*. *Br. J. Exp. Pathol.*  
868 10, 226.
- 869 Fleming, A. (1944). The Discovery of Penicillin. *Br. Med. Bull.* 2, 4–5.  
870 doi:10.1093/oxfordjournals.bmb.a071032.
- 871 Funabashi, M., Funa, N., and Horinouchi, S. (2008). Phenolic lipids synthesized by  
872 type III polyketide synthase confer penicillin resistance on *Streptomyces*  
873 *griseus*. *J. Biol. Chem.* 283, 13983–91. doi:10.1074/jbc.M710461200.
- 874 Gallagher, K. A., Wanger, G., Henderson, J., Llorente, M., Hughes, C. C., and Jensen,  
875 P. R. (2017). Ecological implications of hypoxia-triggered shifts in secondary  
876 metabolism. *Environ. Microbiol.* 19, 2182–2191. doi:10.1111/1462-2920.13700.
- 877 Gaynes, R. (2017). The Discovery of Penicillin—New Insights After More Than 75  
878 Years of Clinical Use. *Emerg. Infect. Dis.* 23, 849–853.  
879 doi:10.3201/eid2305.161556.



- Gerard, J., Haden, P., Kelly, M. T., and Andersen, R. J. (1996). Loloatin B, A cyclic decapeptide antibiotic produced in culture by a tropical marine bacterium. *Tetrahedron Lett.* 37, 7201–7204. doi:10.1016/0040-4039(96)01624-3.
- Geyer, R., Jambeck, J. R., and Law, K. L. (2017). Production, use, and fate of all plastics ever made. *Sci. Adv.* 3, e1700782. doi:10.1126/sciadv.1700782.
- Goodfellow, M., and Fiedler, H.-P. (2010). A guide to successful bioprospecting: informed by actinobacterial systematics. *Antonie Van Leeuwenhoek* 98, 119–142. doi:10.1007/s10482-010-9460-2.
- Haight, T. H., and Finland, M. (1952). The Antibacterial Action of Erythromycin. *Exp. Biol. Med.* 81, 175–183. doi:10.3181/00379727-81-19815.
- Han, X., Liu, W., Huang, J.-W., Ma, J., Zheng, Y., Ko, T.-P., et al. (2017). Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* 8, 2106. doi:10.1038/s41467-017-02255-z.
- Hassan, S. S. ul, Anjum, K., Abbas, S. Q., Akhter, N., Shagufta, B. I., Shah, S. A. A., et al. (2017). Emerging biopharmaceuticals from marine actinobacteria. *Environ. Toxicol. Pharmacol.* 49, 34–47. doi:10.1016/j.etap.2016.11.015.
- Hendlin, D., Stapley, E. O., Jackson, M., Wallick, H., Miller, A. K., Wolf, F. J., et al. (1969). Phosphonomycin, a New Antibiotic Produced by Strains of *Streptomyces*. *Science*. 166, 122–123. doi:10.1126/science.166.3901.122.
- Hodgson, D. A. (2004). Primary metabolism and its control in streptomycetes: A most unusual group of bacteria. 42, 47–238. doi:10.1016/s0065-2911(00)42003-5.
- Hoskisson, P. A., and van Wezel, G. P. (2019). *Streptomyces coelicolor*. *Trends Microbiol.* 27, 468–469. doi:10.1016/j.tim.2018.12.008.
- Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P., and Chatonnet, A. (2004). ESTHER, the database of the  $\alpha/\beta$ -hydrolase fold superfamily of proteins. *Nucleic Acids Res.* 32, 145D – 147. doi:10.1093/nar/gkh141.
- Hsu, K. J., Tseng, M., Don, T. M., and Yang, M. K. (2012). Biodegradation of poly( $\beta$ -hydroxybutyrate) by a novel isolate of *Streptomyces bangladeshensis* 77T-4. *Bot. Stud.* 53, 307–313.
- Hutchinson, C. R. (1997). Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem. Rev.* 97, 2525–2536.
- Indraningrat, A., Smidt, H., and Sipkema, D. (2016). Bioprospecting Sponge-Associated Microbes for Antimicrobial Compounds. *Mar. Drugs* 14, 87. doi:10.3390/md14050087.
- Jackson, S., Crossman, L., Almeida, E., Margassery, L., Kennedy, J., and Dobson, A. (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene

- 916 Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates.  
917 *Mar. Drugs* 16, 67. doi:10.3390/md16020067.
- 918 Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., et al.  
919 (2015). Plastic waste inputs from land into the ocean. *Science*. 347, 768–771.  
920 doi:10.1126/science.1260352.
- 921 Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., et al. (2018). Structural  
922 insight into molecular mechanism of poly(ethylene terephthalate) degradation.  
923 *Nat. Commun.* 9, 382. doi:10.1038/s41467-018-02881-1.
- 924 Katsuyama, Y., and Ohnishi, Y. (2012). “Type III Polyketide Synthases in  
925 Microorganisms,” in *Methods in Enzymology* (Elsevier Inc.), 359–377.  
926 doi:10.1016/B978-0-12-394290-6.00017-3.
- 927 Kawai, F., Kawabata, T., and Oda, M. (2019). Current knowledge on enzymatic PET  
928 degradation and its possible application to waste stream management and  
929 other fields. *Appl. Microbiol. Biotechnol.*, 1–16. doi:10.1007/s00253-019-09717-y.
- 930 Kemung, H. M., Tan, L. T.-H., Khan, T. M., Chan, K.-G., Pusparajah, P., Goh, B.-H.,  
931 et al. (2018). *Streptomyces* as a Prominent Resource of Future Anti-MRSA  
932 Drugs. *Front. Microbiol.* 9, 2221. doi:10.3389/fmicb.2018.02221.
- 933 Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., et al. (2009).  
934 Isolation and Analysis of Bacteria with Antimicrobial Activities from the  
935 Marine Sponge *Haliclona simulans* Collected from Irish Waters. *Mar.*  
936 *Biotechnol.* 11, 384–396. doi:10.1007/s10126-008-9154-1.
- 937 Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007). Structure  
938 and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.*  
939 76, 195–221. doi:10.1146/annurev.biochem.76.053105.093515.
- 940 Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., et al.  
941 (2018). Total Synthesis of the Nonribosomal Peptide Surugamide B and  
942 Identification of a New Offloading Cyclase Family. *Angew. Chemie Int. Ed.* 57,  
943 9447–9451. doi:10.1002/anie.201805541.
- 944 Kwan, J. C., Tianero, M. D. B., Donia, M. S., Wyche, T. P., Bugni, T. S., and Schmidt,  
945 E. W. (2014). Host Control of Symbiont Natural Product Chemistry in Cryptic  
946 Populations of the Tunicate *Lissoclinum patella*. *PLoS One* 9, e95850.  
947 doi:10.1371/journal.pone.0095850.
- 948 Law, J. W., Tan, K., Wong, S. H., Mutalib, N. A., and Lee, L. (2018). Taxonomic and  
949 Characterization Methods of *Streptomyces*: A Review. *Prog. Microbes Mol.*  
950 *Biol.* 1, 1–13.

- 951 Lebreton, L., and Andrady, A. (2019). Future scenarios of global plastic waste  
952 generation and disposal. *Palgrave Commun.* 5, 6. doi:10.1057/s41599-018-0212-  
953 7.
- 954 Lee, N., Hwang, S., Lee, Y., Cho, S., Palsson, B., and Cho, B.-K. (2019). Synthetic  
955 Biology Tools for Novel Secondary Metabolite Discovery in *Streptomyces*. *J.*  
956 *Microbiol. Biotechnol.* 29, 667–686. doi:10.4014/jmb.1904.04015.
- 957 Lewis, K. (2012). Antibiotics: Recover the lost art of drug discovery. *Nature* 485,  
958 439–440. doi:10.1038/485439a.
- 959 Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., et al. (2018). Protein Crystallography  
960 and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate)  
961 Hydrolase PETase from *Ideonella sakaiensis*. *ChemBioChem* 19, 1471–1475.  
962 doi:10.1002/cbic.201800097.
- 963 Liu, C., Shi, C., Zhu, S., Wei, R., and Yin, C.-C. (2019a). Structural and functional  
964 characterization of polyethylene terephthalate hydrolase from *Ideonella*  
965 *sakaiensis*. *Biochem. Biophys. Res. Commun.* 508, 289–294.  
966 doi:10.1016/j.bbrc.2018.11.148.
- 967 Liu, M., El-Hossary, E. M., Oelschlaeger, T. A., Donia, M. S., Quinn, R. J., and  
968 Abdelmohsen, U. R. (2019b). Potential of marine natural products against  
969 drug-resistant bacterial infections. *Lancet Infect. Dis.* 19, e237–e245.  
970 doi:10.1016/S1473-3099(18)30711-4.
- 971 Luckner, M. (2014). *Secondary Metabolism in Microorganisms, Plants and Animals*.  
972 Springer Berlin Heidelberg.
- 973 Manivasagan, P., Kang, K.-H., Sivakumar, K., Li-Chan, E. C. Y., Oh, H.-M., and Kim,  
974 S.-K. (2014). Marine actinobacteria: An important source of bioactive natural  
975 products. *Environ. Toxicol. Pharmacol.* 38, 172–188.  
976 doi:10.1016/j.etap.2014.05.014.
- 977 Martínez-Núñez, M. A., and López, V. E. L. y (2016). Nonribosomal peptides  
978 synthetases and their applications in industry. *Sustain. Chem. Process.* 4, 1–8.  
979 doi:10.1186/s40508-016-0057-6.
- 980 Mateos, M. V., Blade, J., Prosper, F., Lahuerta, J. J., Garcia-Sanz, R., Cibeira, M., et al.  
981 (2005). Preliminary Report from an Exploratory Phase II Trial with Plitidepsin  
982 (Aplidin®) in Patients with Refractory/Relapsed Multiple Myeloma. *Blood* 106.
- 983 Matsuda, K., Kobayashi, M., Kuranaga, T., Takada, K., Ikeda, H., Matsunaga, S., et  
984 al. (2019a). SurE is a trans-acting thioesterase cyclizing two distinct non-  
985 ribosomal peptides. *Org. Biomol. Chem.* 17, 1058–1061.  
986 doi:10.1039/C8OB02867B.

- 987 Matsuda, K., Kuranaga, T., Sano, A., Ninomiya, A., Takada, K., and Wakimoto, T.  
 988 (2019b). The Revised Structure of the Cyclic Octapeptide Surugamide A. *Chem.*  
 989 *Pharm. Bull.* 67, 476–480. doi:10.1248/cpb.c19-00002.
- 990 McCormick, J. R., and Flärdh, K. (2012). Signals and regulators that govern  
 991 *Streptomyces* development. *FEMS Microbiol. Rev.* 36, 206–31.  
 992 doi:10.1111/j.1574-6976.2011.00317.x.
- 993 McCormick, M. H., McGuire, J. M., Pittenger, G. E., Pittenger, R. C., and Stark, W.  
 994 M. (1955). Vancomycin, a new antibiotic. I. Chemical and biologic properties.  
 995 *Antibiot. Annu.* 3, 606–11.
- 996 Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et  
 997 al. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem.*  
 998 *Biol.* 11, 625–631. doi:10.1038/nchembio.1890.
- 999 Mehbub, M. F., Tanner, J. E., Barnett, S. J., Franco, C. M. M., and Zhang, W. (2016).  
 1000 The role of sponge-bacteria interactions: the sponge *Aplysilla rosea* challenged  
 1001 by its associated bacterium *Streptomyces* ACT-52A in a controlled aquarium  
 1002 system. *Appl. Microbiol. Biotechnol.* 100, 10609–10626. doi:10.1007/s00253-016-  
 1003 7878-9.
- 1004 Miller, B. R., and Gulick, A. M. (2016). “Structural biology of nonribosomal peptide  
 1005 synthetases,” in *Nonribosomal Peptide and Polyketide Biosynthesis* (Humana  
 1006 Press, New York, NY), 3-29. doi:10.1007/978-1-4939-3375-4.
- 1007 Mohr, K. I. (2016). “History of Antibiotics Research,” in *How to Overcome the*  
 1008 *Antibiotic Crisis* (Springer, Cham), 237–272. doi:10.1007/82\_2016\_499.
- 1009 Muto, A., and Osawa, S. (1987). The guanine and cytosine content of genomic DNA  
 1010 and bacterial evolution (biased mutation pressure/codon usage/neutral theory).  
 1011 *Proc. Natl. Acad. Sci. USA* 84, 166–169.
- 1012 Nardini, M., and Dijkstra, B. W. (1999).  $\alpha/\beta$  Hydrolase fold enzymes: the family  
 1013 keeps growing. *Curr. Opin. Struct. Biol.* 9, 732–737. doi:10.1016/S0959-  
 1014 440X(99)00037-8.
- 1015 Navarro-Muñoz, J. C., Selem-Mojica, N., Mallowney, M. W., Kautsar, S., Tryon, J. H.,  
 1016 Parkinson, E. I., et al. (2018). A computational framework for systematic  
 1017 exploration of biosynthetic diversity from large-scale genomic data. *bioRxiv*,  
 1018 445270. doi:10.1101/445270.
- 1019 Nesme, J., Achouak, W., Agathos, S. N., Bailey, M., Baldrian, P., Brunel, D., et al.  
 1020 (2016) Back to the Future of Soil Metagenomics. *Front. Microbiol.* 7:73. doi:  
 1021 10.3389/fmicb.2016.00073.

- 1022 Newman, D. J., and Cragg, G. M. (2016). Natural Products as Sources of New  
 1023 Drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661.  
 1024 doi:10.1021/acs.jnatprod.5b01055.
- 1025 Newman, D. J., Cragg, G. M., and Snader, K. M. (2000). The influence of natural  
 1026 products upon drug discovery. *Nat. Prod. Rep.* 17, 215–234.  
 1027 doi:10.1039/a902202c.
- 1028 Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., et al.  
 1029 (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an  
 1030 Unrelated Decapeptide, Surugamide F. *ChemBioChem* 17, 1709–1712.  
 1031 doi:10.1002/cbic.201600350.
- 1032 Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., et al.  
 1033 (1992). The  $\alpha/\beta$  hydrolase fold. *Protein Eng. Des. Sel.* 5, 197–211.  
 1034 doi:10.1093/protein/5.3.197.
- 1035 O'Neill, J. I. M. (2014). Antimicrobial resistance: tackling a crisis for the health and  
 1036 wealth of nations. *Rev. Antimicrob. Resist.* 20, 1–16.
- 1037 Pan, R., Bai, X., Chen, J., Zhang, H., and Wang, H. (2019). Exploring Structural  
 1038 Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A  
 1039 Literature Review. *Front. Microbiol.* 10, 294. doi:10.3389/fmicb.2019.00294.
- 1040 Pendleton, J. N., Gorman, S. P., and Gilmore, B. F. (2013). Clinical relevance of the  
 1041 ESKAPE pathogens. *Expert Rev. Anti. Infect. Ther.* 11, 297–308.  
 1042 doi:10.1586/eri.13.12.
- 1043 PlasticsEurope (2018). Plastics – the Facts 2018. *PlasticsEurope*. Available at:  
 1044 [https://www.plasticseurope.org/en/resources/publications/619-plastics-facts-](https://www.plasticseurope.org/en/resources/publications/619-plastics-facts-2018)  
 1045 2018 [Accessed May 14, 2019].
- 1046 Putnam, L. E., Hendricks, F. D., and Welch, H. (1953). Tetracycline, a new antibiotic.  
 1047 *Antibiot. Chemother. (Northfield, Ill.)* 3, 1183–6.
- 1048 Ragaert, K., Delva, L., and Van Geem, K. (2017). Mechanical and chemical recycling  
 1049 of solid plastic waste. *Waste Manag.* 69, 24–58.  
 1050 doi:10.1016/J.WASMAN.2017.07.044.
- 1051 Raja, A., LaBonte, J., Lebbos, J., and Kirkpatrick, P. (2003). Daptomycin. *Nat. Rev.*  
 1052 *Drug Discov.* 2, 943–944. doi:10.1038/nrd1258.
- 1053 Reimer, J. M., Haque, A. S., Tarry, M. J., and Schmeing, T. M. (2018). Piecing  
 1054 together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.* 49, 104–113.  
 1055 doi:10.1016/j.sbi.2018.01.011.

- Robbins, T., Liu, Y.-C., Cane, D. E., and Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol.* 41, 10–18. doi:10.1016/j.sbi.2016.05.009.
- Romano, S., Jackson, S., Patry, S., and Dobson, A. (2018). Extending the “One Strain Many Compounds” (OSMAC) Principle to Marine Microorganisms. *Mar. Drugs* 16, 244. doi:10.3390/md16070244.
- Santillo, D., Miller, K., and Johnston, P. (2017). Microplastics as contaminants in commercially important seafood species. *Integr. Environ. Assess. Manag.* 13, 516–521. doi:10.1002/ieam.1909.
- Sathiyarayanan, G., Gandhimathi, R., Sabarathnam, B., Seghal Kiran, G., and Selvin, J. (2014). Optimization and production of pyrrolidone antimicrobial agent from marine sponge-associated *Streptomyces* sp. MAPS15. *Bioprocess Biosyst. Eng.* 37, 561–573. doi:10.1007/s00449-013-1023-2.
- Schaffer, L., Finkelstein, J., Hohn, A., and Djerassi, I. (1963). Lincomycin—a New Antibiotic. *Clin. Pediatr. (Phila)*. 2, 642–645. doi:10.1177/000992286300201110.
- Schatz, A., Bugle, E., and Waksman, S. A. (1944). Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.\*. *Exp. Biol. Med.* 55, 66–69. doi:10.3181/00379727-55-14461.
- Schinke, C., Martins, T., Queiroz, S. C. N., Melo, I. S., and Reyes, F. G. R. (2017). Antibacterial Compounds from Marine Bacteria, 2010–2015. *J. Nat. Prod.* 80, 1215–1228. doi:10.1021/acs.jnatprod.6b00235.
- Schneemann, I., Kajahn, I., Ohlendorf, B., Zinecker, H., Erhard, A., Nagel, K., et al. (2010). Mayamycin, a Cytotoxic Polyketide from a *Streptomyces* Strain Isolated from the Marine Sponge *Halichondria panicea*. *J. Nat. Prod.* 73, 1309–1312. doi:10.1021/np100135b.
- Schrör, K. (2016). *Acetylsalicylic acid*. John Wiley & Sons.
- Sélem-Mojica, N., Aguilar, C., Gutiérrez-García, K., Martínez-Guerrero, C. E., and Barona-Gómez, F. (2018). EvoMining reveals the origin and fate of natural products biosynthetic enzymes. *bioRxiv*, 482273. doi:10.1101/482273.
- Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* 7, 285–295. doi:10.1016/S1367-5931(03)00020-6.
- Shimizu, Y., Ogata, H., and Goto, S. (2017). Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem* 18, 50–65. doi:10.1002/cbic.201600522.

- 1091 Silva, C., Da, S., Silva, N., Matamá, T., Araújo, R., Martins, M., et al. (2011).  
 1092 Engineered *Thermobifida fusca* cutinase with increased activity on polyester  
 1093 substrates. *Biotechnol. J.* 6, 1230–1239. doi:10.1002/biot.201000391.
- 1094 Sirijan, S., and Nitaya, I. (2016). Mechanisms of antimicrobial resistance in bacteria  
 1095 in ESKAPE pathogens. *Biomed Res. Int.* 2016, 1–8. doi:10.1016/j.ajic.2006.05.219.
- 1096 Staunton, J., and Wilkinson, B. (1997). Biosynthesis of Erythromycin and  
 1097 Rapamycin. *Chem. Rev.* 97, 2611–2630.
- 1098 Sujatha, P., Bapi Raju, K. V. V. S. N., and Ramana, T. (2005). Studies on a new  
 1099 marine streptomycete BT-408 producing polyketide antibiotic SBR-22 effective  
 1100 against methicillin resistant *Staphylococcus aureus*. *Microbiol. Res.* 160, 119–  
 1101 126. doi:10.1016/j.micres.2004.10.006.
- 1102 Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., et al. (2013).  
 1103 Surugamides A–E, Cyclic Octapeptides with Four d-Amino Acid Residues,  
 1104 from a Marine *Streptomyces* sp.: LC–MS-Aided Inspection of Partial  
 1105 Hydrolysates for the Distinction of d- and l-Amino Acid Residues in the  
 1106 Sequence. *J. Org. Chem.* 78, 6746–6750. doi:10.1021/jo400708u.
- 1107 Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M. E., and Seipke, R. F.  
 1108 (2019). A *trans*-Acting Cyclase Offloading Strategy for Nonribosomal Peptide  
 1109 Synthetases. *ACS Chem. Biol.* 14, 845–849. doi:10.1021/acscchembio.9b00095.
- 1110 Tokiwa, Y., and Calabia, B. P. (2004). Review Degradation of microbial polyesters.  
 1111 *Biotechnol. Lett.* 26, 1181–1189. doi:10.1023/B:BILE.0000036599.15302.e5.
- 1112 Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I., and Miller, A. A. (2015).  
 1113 ESKAPEing the labyrinth of antibacterial discovery. *Nat. Rev. Drug Discov.* 14,  
 1114 529–542. doi:10.1038/nrd4572.
- 1115 Torsvik, V., Goksøyr J., and Daae F. L. (1990). High diversity in DNA of soil bacteria.  
 1116 *Appl. Environ. Microbiol.* 56(3), 782–787.
- 1117 van der Meij, A., Worsley, S. F., Hutchings, M. I., and van Wezel, G. P. (2017).  
 1118 Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol.*  
 1119 *Rev.* 41, 392–416. doi:10.1093/femsre/fux005.
- 1120 Vavra, J. J., Deboer, C., Dietz, A., Hanka, L. J., and Sokolski, W. T. (1959).  
 1121 Streptozotocin, a new antibacterial antibiotic. *Antibiot. Annu.* 7, 230–5.
- 1122 Wade, W. (2002). Unculturable bacteria—the uncharacterized organisms that cause  
 1123 oral infections. *Journal of the Royal Society of Medicine.* 95(2), 81–83.  
 1124 doi:10.1258/jrsm.95.2.81.

- 1125 Waksman, S. A., and Lechevalier, H. A. (1949). Neomycin, a New Antibiotic Active  
1126 against Streptomycin-Resistant Bacteria, including Tuberculosis Organisms.  
1127 *Science*. 109, 305–307. doi:10.1126/science.109.2830.305.
- 1128 Waksman, S. A., and Woodruff, H. B. (1942). Streptothricin, a New Selective  
1129 Bacteriostatic and Bactericidal Agent, Particularly Active Against Gram-  
1130 Negative Bacteria. *Exp. Biol. Med.* 49, 207–210. doi:10.3181/00379727-49-13515.
- 1131 Watve, M., Tickoo, R., Jog, M., and Bhole, B. (2001). How many antibiotics are  
1132 produced by the genus *Streptomyces*? *Arch. Microbiol.* 176, 386–390.  
1133 doi:10.1007/s002030100345.
- 1134 Weber, J., Leung, J., Swanson, S., Idler, K., and McAlpine, J. (1991). An  
1135 erythromycin derivative produced by targeted gene disruption in  
1136 *Saccharopolyspora erythraea*. *Science*. 252, 114–117.  
1137 doi:10.1126/science.2011746.
- 1138 Weber, T., and Kim, H. U. (2016). The secondary metabolite bioinformatics portal:  
1139 Computational tools to facilitate synthetic biology of secondary metabolite  
1140 production. *Synth. Syst. Biotechnol.* doi:10.1016/j.synbio.2015.12.002.
- 1141 Wei, R., and Zimmermann, W. (2017a). Biocatalysis as a green route for recycling  
1142 the recalcitrant plastic polyethylene terephthalate. *Microb. Biotechnol.* 10,  
1143 1302–1307. doi:10.1111/1751-7915.12714.
- 1144 Wei, R., and Zimmermann, W. (2017b). Microbial enzymes for the recycling of  
1145 recalcitrant petroleum-based plastics: how far are we? *Microb. Biotechnol.* 10,  
1146 1308–1322. doi:10.1111/1751-7915.12710.
- 1147 Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al.  
1148 (2018). DrugBank 5.0: a major update to the DrugBank database for 2018.  
1149 *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037.
- 1150 Wu, B., Oesker, V., Wiese, J., Malien, S., Schmaljohann, R., Imhoff, J., et al. (2014).  
1151 Spirocyclic Drimanes from the Marine Fungus *Stachybotrys* sp. Strain MF347.  
1152 *Mar. Drugs* 12, 1924–1938. doi:10.3390/md12041924.
- 1153 Wu, Q., Zhang, G., Wang, B., Li, X., Yue, S., Chen, J., et al. (2018). Production and  
1154 Identification of Inthomycin B Produced by a Deep-Sea Sediment-Derived  
1155 *Streptomyces* sp. YB104 Based on Cultivation-Dependent Approach. *Curr.*  
1156 *Microbiol.* 75, 942–951. doi:10.1007/s00284-018-1469-1.
- 1157 Xu, F., Nazari, B., Moon, K., Bushin, L. B., and Seyedsayamdost, M. R. (2017).  
1158 Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074  
1159 Using High-Throughput Elicitor Screens. *J. Am. Chem. Soc.* 139, 9203–9212.  
1160 doi:10.1021/jacs.7b02716.



- 1161 Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., et al.  
1162 (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate).  
1163 *Science*. 351, 1196–1199. doi:10.1126/science.aad6359.
- 1164 Yu, D., Xu, F., Zeng, J., and Zhan, J. (2012). Type III polyketide synthases in natural  
1165 product biosynthesis. *IUBMB Life* 64, 285–295. doi:10.1002/iub.1005.
- 1166 Zhang, Z., and Wood, W. I. (2003). A profile hidden Markov model for signal  
1167 peptides generated by HMMER. *Bioinformatics* 19, 307–308.  
1168 doi:10.1093/bioinformatics/19.2.307.
- 1169 Zhou, Y., Lin, X., Xu, C., Shen, Y., Wang, S.-P., Liao, H., et al. (2019). Investigation of  
1170 Penicillin Binding Protein (PBP)-like Peptide Cyclase and Hydrolase in  
1171 Surugamide Non-ribosomal Peptide Biosynthesis. *Cell Chem. Biol.* 26, 737-  
1172 744.e4. doi:10.1016/j.chembiol.2019.02.010.
- 1173 Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in  
1174 microbes – a review. *Nat. Prod. Rep.* 33, 988–1005. doi:10.1039/C6NP00025H.

# Chapter 2

## Comparative genomics of marine sponge-derived *Streptomyces* spp. isolates SM17 and SM18 with their closest terrestrial relatives provides novel insights into environmental niche adaptations and secondary metabolite biosynthesis potential

Eduardo L. Almeida<sup>1</sup>, Andrés Felipe Carrillo Rincón<sup>1</sup>, Stephen A. Jackson<sup>1,2</sup> and Alan D.W. Dobson<sup>1,2</sup>

<sup>1</sup>School of Microbiology, University College Cork, Cork, Ireland

<sup>2</sup>Environmental Research Institute, University College Cork, Cork, Ireland

The work presented in this chapter has been published in the journal *Frontiers in Microbiology*, 26 July 2019, <https://doi.org/10.3389/fmicb.2019.01713>.

**Citation:** Almeida E. L., Rincón A. F. C., Jackson S. A. and Dobson A. D. W. (2019). Comparative Genomics of Marine Sponge-Derived *Streptomyces* spp. Isolates SM17 and SM18 With Their Closest Terrestrial Relatives Provides Novel Insights Into Environmental Niche Adaptations and Secondary Metabolite Biosynthesis Potential. *Front. Microbiol.* 10:1713. doi: 10.3389/fmicb.2019.01713.

1175    **1.    Abstract**

1176            The emergence of antibiotic resistant microorganisms has led to an increased  
1177 need for the discovery and development of novel antimicrobial compounds. Frequent  
1178 rediscovery of the same natural products (NPs) continues to decrease the likelihood  
1179 of the discovery of new compounds from soil bacteria. Thus, efforts have shifted  
1180 toward investigating microorganisms and their secondary metabolite biosynthesis  
1181 potential, from diverse niche environments, such as those isolated from marine  
1182 sponges. Here we investigated at the genomic level two *Streptomyces* spp. strains,  
1183 namely SM17 and SM18, isolated from the marine sponge *Haliclona simulans*, with  
1184 previously reported antimicrobial activity against clinically relevant pathogens; using  
1185 single molecule real-time (SMRT) sequencing. We performed a series of comparative  
1186 genomic analyses on SM17 and SM18 with their closest terrestrial relatives, namely *S.*  
1187 *albus* J1074 and *S. pratensis* ATCC 33331 respectively; in an effort to provide further  
1188 insights into potential environmental niche adaptations (ENAs) of marine sponge-  
1189 associated *Streptomyces*, and on how these adaptations might be linked to their  
1190 secondary metabolite biosynthesis potential. Prediction of secondary metabolite  
1191 biosynthetic gene clusters (smBGCs) indicated that, even though the marine isolates  
1192 are closely related to their terrestrial counterparts at a genomic level; they potentially  
1193 produce different compounds. SM17 and SM18 displayed a better ability to grow in  
1194 high salinity medium when compared to their terrestrial counterparts, and further  
1195 analysis of their genomes indicated that they possess a pool of 29 potential ENA  
1196 genes that are absent in *S. albus* J1074 and *S. pratensis* ATCC 33331. This ENA gene  
1197 pool included functional categories of genes that are likely to be related to niche  
1198 adaptations and which could be grouped based on potential biological functions such  
1199 as osmotic stress defense; transcriptional regulation; symbiotic interactions;  
1200 antimicrobial compound production and resistance; ABC transporters; together with  
1201 horizontal gene transfer and defense-related features.

## 1202 2. Introduction

1203 With the emergence and rapid spread of antibiotic resistant microorganisms,  
1204 displaying resistance to many currently available antibiotics, a concerted effort  
1205 continues to be needed to discover novel antimicrobial agents (Thabit et al., 2015;  
1206 Rolain et al., 2016). Members of the *Streptomyces* genus are also known to produce  
1207 a broad range of other natural products (NPs) which possess immunosuppressant,  
1208 anti-fungal, anti-cancer, anti-parasitic and anti-thrombotic activities (Hwang et al.,  
1209 2014; Ser et al., 2017). However, the frequent re-discovery of previously  
1210 characterized bioactive compounds from terrestrial *Streptomyces*, has somewhat  
1211 limited the interest of researchers in terrestrial ecosystems as potential reservoirs  
1212 for novel biomolecules (Yagüe et al., 2012; Dalisay et al., 2013; Paulus et al., 2017).  
1213 Instead, interest has begun to focus on the isolation of *Streptomyces* from other  
1214 environmental niches; with *Streptomyces* involved in symbiotic relationships or  
1215 associated with plants, insects, fungi, lichens, sea-cucumbers, seaweeds and marine  
1216 sponges also attracting increased attention as potential reservoirs for these types of  
1217 bioactive molecules (Motohashi et al., 2010; Seipke et al., 2012; van der Meij et al.,  
1218 2017). The ability of these *Streptomyces* to colonize such a wide variety of hosts is  
1219 due in part to their ability to produce useful NPs, such as antimicrobials which  
1220 help their hosts defend themselves against predators or pathogenic bacteria and  
1221 fungi (Adnani et al., 2017; van der Meij et al., 2017).

1222 Marine ecosystems are attracting particular attention, where extreme and  
1223 rapidly changing environmental conditions such as differences in pressure, salinity,  
1224 pH, light intensity, temperature and oligotrophic conditions are believed to be  
1225 linked to secondary metabolites production (Abdelmohsen et al., 2014; van der  
1226 Meij et al., 2017). In this respect, marine ecosystems have been a particularly  
1227 fruitful source of *Streptomyces* strains which have the potential to produce new  
1228 bioactive NPs (Hassan et al., 2017; Jin et al., 2018; Xu et al., 2018), with marine

1229 *Streptomyces* being isolated from seashores, coastal waters, bottom sediments,  
 1230 fishes, molluscs, sponges, seaweeds and mangroves (Manivasagan et al., 2014; Ser  
 1231 et al., 2017).

1232 Marine sponges (phylum *Porifera*) in particular are known to be a rich  
 1233 source of bioactive compounds, many of which are produced by the bacteria which  
 1234 reside within the sponge host (Abdelmohsen et al., 2014; Fuerst, 2014). Many of  
 1235 these bioactives have antimicrobial activities, making these sponge-associated  
 1236 microbial and fungal communities a potentially valuable source of novel  
 1237 antimicrobials (Baker et al., 2009; Flemer et al., 2012; Hoppers et al., 2015;  
 1238 Indraningrat et al., 2016; Jackson et al., 2018). While sponge bacteria-derived  
 1239 antimicrobial compounds have to date been identified from 35 different genera, the  
 1240 most predominant producing genera include *Streptomyces*, *Pseudovibrio* and  
 1241 *Bacillus* strains (Indraningrat et al., 2016). Of these, *Streptomyces* are the  
 1242 predominant genus, producing around 30% of the compounds identified to date  
 1243 (Indraningrat et al., 2016). Good examples of bioactive compounds produced from  
 1244 *Streptomyces* associated with marine sponges include: mayamycin, produced by  
 1245 *Streptomyces* sp. HB202 isolated from *Halichondria panicea* (Schneemann et al.,  
 1246 2010); the naphthacene glycoside SF2446A2, produced by *Streptomyces* sp. RV15  
 1247 isolated from *Dysidea tupha* (Reimer et al., 2015); and Petrocidin A, produced by  
 1248 *Streptomyces* sp. SBT348 isolated from *Petrosia ficiformis* (Cheng et al., 2017).

1249 As previously mentioned, in addition to marine sponges, many  
 1250 *Streptomyces* strains have also evolved symbiotic relationships with plants, fungi,  
 1251 and insects, amongst others; and there is increasing evidence that the host may  
 1252 control which metabolic pathways are activated within their symbionts, such as in  
 1253 the tunicate *Lissoclinum patella* and the squid *Euprymna scolopes* (Kwan et al.,  
 1254 2014; Gromek et al., 2016). In *Streptomyces* spp., it is clear that not only do they  
 1255 benefit from the resources of the hosts they interact with, but that these interactions

control the expression of secondary metabolite biosynthetic gene clusters (smBGCs); thereby promoting the high degree of chemical diversity observed in the secondary metabolites being produced by these organisms (van der Meij et al., 2017). An example is the recent report that exposure of the endosymbiont *Streptomyces* ACT-52A to *Aplysilla rosea* promoted production of bioactive compounds with antibacterial activity (Mehbub et al., 2016). The factors involved in controlling the expression of these smBGCs are likely to be quite diverse, given the large degree of variability in the habitats and potential hosts, and how they are presumably influencing the secondary metabolite biosynthetic potential of *Streptomyces* symbionts (Adnani et al., 2017). Thus, it is clear that an increased knowledge of the genetics underpinning the interactions and signaling between the sponge host and the symbiont is required, through identification of smBGCs in the genomes of these sponge associated *Streptomyces* strains, coupled with identification of potential environmental “triggers” from the sponge, from other sponge endosymbionts, and/or from the surrounding marine environment that may regulate transcription of these smBGCs (Mehbub et al., 2016; Adnani et al., 2017; van der Meij et al., 2017).

To this end, we recently sequenced the genomes of 13 *Streptomyces* spp. isolated from both shallow water and deep-sea sponges, that displayed antimicrobial activities against a number of clinically relevant bacterial and yeast species (Kennedy et al., 2009; Jackson et al., 2018). Using the antiSMASH (antibiotics and Secondary Metabolite Analysis Shell) software (Blin et al., 2017), the strains were found to host abundant smBGCs which potentially encode polyketides, non-ribosomal peptide synthases (NRPS), siderophores, lantipeptides, and bacteriocins (Jackson et al., 2018). Thus, these strains appear to be a promising source of novel bioactive secondary metabolites, as the abundance and diversity of smBGCs displayed high degrees of novelty. In addition, the strains were enriched for genes potentially involved in the biosynthesis and transport of compatible solutes and for heat-shock proteins, genes

1283 which are typically associated with marine adaptations (Penn and Jensen, 2012; Tian  
1284 et al., 2016).

1285         Around sixty marine adaptation genes (MAGs) have previously been  
1286 proposed for the obligate marine actinomycete genus *Salinispora*, with the function  
1287 of these genes being associated with electron transport, sodium and ABC  
1288 transporters, together with channels and pores (Penn and Jensen, 2012). Even though  
1289 sponge-associated *Streptomyces* are marine bacteria, the environmental niche  
1290 occupied by these organisms differs quite markedly from *Salinispora*, thus the  
1291 genetic adaptations may not necessarily be similar. This was confirmed by the Zotchev  
1292 group, when the draft genome of two sponge associated *Streptomyces* strains were  
1293 analyzed for MAGs, revealing the presence of only seven of the *Salinispora* MAG  
1294 gene pool (Ian et al., 2014). They suggested that specific marine sponge genetic  
1295 adaptations may exist, given that different genes were identified in these sponge-  
1296 associated *Streptomyces* which were absent in their soil counterparts (Ian et al., 2014).  
1297 However, drawing conclusions for these genetic adaptations is quite difficult due to  
1298 the limited number of sponge-associated *Streptomyces* genomes that are currently  
1299 available. To this end, we sequenced the genomes of *Streptomyces* strains SM17 and  
1300 SM18, two of the aforementioned 13 sponge-derived *Streptomyces* spp. that had  
1301 displayed antimicrobial activity, using the PacBio RSII Single Molecule, Real-Time  
1302 (SMRT) sequencing platform. This allowed us to study not only the smBGCs that  
1303 these bacteria possess, but also other genetic characteristics that may be involved in  
1304 their life cycle; such as for example adaptation to the marine environment and  
1305 symbiosis. By employing comparative genomics, we compared the genomes of these  
1306 strains with their most closely related terrestrial type-strain relatives, with complete  
1307 genomes available in the GenBank database (namely *S. albus* J1074 for SM17 and *S.*  
1308 *pratensis* ATCC 33331 for SM18), in an attempt to identify genes potentially  
1309 associated with ENA, together with genes encoding potentially novel smBGCs.

### 1310 3. Material and methods

#### 1311 3.1. Bacterial strains, maintenance and differential growth assessment

1312 The SM17 and SM18 strains were isolated from the marine sponge *Haliclona*  
1313 *simulans* (Kilkieran Bay, Galway, Ireland), as previously described (Kennedy et al.,  
1314 2009). The *S. albus* J1074 strain was provided by Dr. Andriy Luzhetskyy  
1315 (Helmholtz Institute for Pharmaceutical Research Saarland, Germany), while *S.*  
1316 *flavogriseus*/*S. pratensis* ATCC 33331 was obtained from the American Type  
1317 Culture Collection (ATCC Inc., United States). SM17, SM18, *S. albus* J1074 and *S.*  
1318 *flavogriseus*/*S. pratensis* ATCC 33331 spores were propagated on mannitol-soya  
1319 (MS) agar medium at 28°C for 8–10 days and stored in 20% glycerol at –80°C.  
1320 Strains were cultivated on ISP2 and ISP2 plus artificial sea water (ASW) medium  
1321 when indicated, for differential growth analysis. The ASW was obtained by adding  
1322 3% Instant Ocean® Sea Salt (Instant Ocean Inc., United States) to the medium. It is  
1323 important to note that the ATCC 33331 strain, due to a more recent taxonomy  
1324 classification (Rong et al., 2013), is described with two different names: in GenBank  
1325 as *S. pratensis* ATCC 33331 (new classification), and in the ATCC® culture collection  
1326 as *S. flavogriseus* ATCC 33331 (old classification). From now on, the ATCC 33331  
1327 isolate will be referred to as *S. pratensis* ATCC 33331.

#### 1328 3.2. Genome sequencing, assembly and annotation

1329 Biomass from the SM17 and SM18 strains was obtained after cultivation on  
1330 TSB medium for 3 days at 28°C and 220 rpm. Genomic DNA from SM17 was  
1331 isolated using the DNeasy Blood & Cell Culture DNA Midi Kit (Qiagen Inc.); and  
1332 by using the phenol-chloroform-isoamyl alcohol extraction method for SM18  
1333 (Wilson, 2001). Genome sequencing was performed by Macrogen (Seoul, South  
1334 Korea), using the PacBio RSII sequencing platform.



1335           The PacBio raw reads were processed and quality filtered using the  
 1336 BamTools toolkit v2.4.1 (subread length >1000, subread quality >0.75) (Barnett et al.,  
 1337 2011). The genome assemblies were performed using the Canu v1.7 software  
 1338 (Koren et al., 2017), followed by assembly polishing using Quiver v2.1.0 (Pacific  
 1339 Biosciences Inc). The assembly coverage check was performed using the BBMap  
 1340 program v37.90 (available at <https://sourceforge.net/projects/bbmap/>). Genome  
 1341 assembly statistics were calculated using the QUAST v4.6.3 program (Gurevich et  
 1342 al., 2013). Genome annotation was performed using the Prokka v1.12 program for  
 1343 this study's analyses (Seemann, 2014), and with the NCBI Prokaryotic Genome  
 1344 Annotation Pipeline for data submission on the GenBank database (Tatusova et al.,  
 1345 2016; Benson et al., 2018). Prediction of smBGCs was performed using the  
 1346 antiSMASH 4 software (Blin et al., 2017). Similarity clustering of smBGCs families  
 1347 was performed using the Biosynthetic Genes Similarity Clustering and Prospecting  
 1348 Engine (BiG-SCAPE, version 2018100) (Navarro-Muñoz et al., 2018) and Cytoscape  
 1349 (v3.7.1) (Shannon et al., 2003), with annotations based on the Minimum  
 1350 Information about a Biosynthetic Gene cluster (MIBiG) repository (v1.4) (Medema  
 1351 et al., 2015). Genome maps were generated using the Artemis v17.0.1 and the  
 1352 DNAPlotter v17.0.1 programs (Rutherford et al., 2000; Carver et al., 2009). Proteins  
 1353 of interest were manually annotated using the NCBI BLAST tool; the GenBank  
 1354 database; and the Conserved Domain Database (CDD) (Johnson et al., 2008;  
 1355 Marchler-Bauer et al., 2015; Benson et al., 2018).

### 1356   **3.3. Comparative genomics**

1357           The closest reference strains for the sponge-derived isolates SM17 and SM18  
 1358 were determined by employing a phylogenetic analysis performed in two steps: (1)  
 1359 based on the 16S rRNA sequence of the SM17 and SM18 isolates, we picked the top  
 1360 30 most similar *Streptomyces* species to each of the isolates (for a total of 60  
 1361 genomes from the database), with complete genome available in GenBank (Benson

et al., 2018), using the NCBI BLAST tool (Johnson et al., 2008) (2) we then performed a phylogenetic analysis employing concatenated sequences (Gadagkar et al., 2005) of the 16S rRNA and the housekeeping genes *atpD* (ATP synthase subunit beta), *gyrB* (DNA gyrase subunit B), *recA* (recombinase RecA), *rpoB* (DNA-directed RNA polymerase subunit beta), and *trpB* (tryptophan synthase beta chain), of the SM17 and SM18 strains, plus the previously determined top 60 most similar *Streptomyces* species. Alignment of the concatenated sequences was performed using the MAFFT program (Katoh and Standley, 2013), and phylogeny was determined using the MrBayes program (Ronquist et al., 2012), applying the General Time Reversible (GTR) model of nucleotide substitution with gamma-distributed rates across sites with a proportion of invariable sites (Waddell and Steel, 1997), and an average standard deviation of split frequencies cut off of 0.01. The final condensed tree, with a posterior probability cut off of 95%, was generated using MEGA X (Kumar et al., 2018) and Inkscape (available at <https://inkscape.org>). To further support genomic similarities between the SM17 and SM18 strain and their closest type-strain terrestrial relative determined with the phylogeny analysis, alignments of the individual housekeeping genes were performed and sequence similarity was determined, using the NCBI BLAST tool (Johnson et al., 2008); and whole genome nucleotide alignments were performed using the MUMmer 3.0 program (Kurtz et al., 2004). Plasmids sequences were determined by similarity searches in the GenBank database (Benson et al., 2018). Orthologous gene analysis was performed using the Roary v3.12.0 program, with an identity cut-off set to 50% (Page et al., 2015). The Roary outputs were processed using the R software environment in the RStudio IDE (Racine, 2012; RStudio Team, 2015; R Core Team, 2018), with data frame handling using the plyr package (Wickham, 2011); and Venn diagrams generated using the venn package (Dusa, 2018).

1388    **3.4. Accession numbers**

1389            The complete genome sequences of SM17, SM18, and the SM17 plasmid  
1390 sequences pSM17A, pSM17B, pSM17C, have been deposited in GenBank under the  
1391 accession numbers CP029338, CP029342, CP029339, CP029340, and CP029341,  
1392 respectively. The closest reference genomes used in this study for comparative  
1393 purposes were *S. albus* J1074 (accession no. CP004370.1) and *S. pratensis* ATCC  
1394 33331 (accession no. CP002475.1).

## 1395 4. Results and discussion

### 1396 4.1. Genome sequencing and assembly

1397 The genomes of the marine sponge-derived *Streptomyces* spp. isolates SM17  
 1398 and SM18 were sequenced using the PacBio RSII sequencing platform, which  
 1399 generated a total of 140,538 and 87,756 subreads respectively, after adapter removal  
 1400 and quality/length filtering (Table 1A). The PacBio sequencing provided long read  
 1401 lengths, averaging 9,702 and 8,923 bp for SM17 and SM18, respectively. Combining  
 1402 the large number of reads and their long length, an approximate sequencing  
 1403 coverage of 194× and 101× was obtained for SM17 and SM18, respectively.

	SM17	SM18
Genome size (bp)	7,179,914*	7,703,166
Number of subreads	140,538	87,756
Average subread length (bp)	9,702	8,923
Approximate average sequencing coverage (fold)	194	101
GC content (%)	73.35	71.84
Number of contigs	4**	1
N50	6,975,788	7,703,166
L50	1	1
Number of coding sequences	6,181	6,670
Number of rRNAs	21	18
Number of tRNAs	78	82
Number of tmRNAs	1	1

1404 **Table 1A:** General characteristics of the SM17 and the SM18 genomes.  
 1405 For SM17, the statistics include the sequence of the chromosome in  
 1406 addition to the sequences determined to represent three plasmids,  
 1407 hence the \* > 7 Mb genome size and the \*\*total of 4 contigs. No  
 1408 plasmids were identified in the SM18 strain, so the statistics  
 1409 represented above are for the chromosome sequence. No gaps or  
 1410 ambiguous bases (Ns) are present in the final genome assemblies.

1411           The genome assemblies for both isolates were of a very high quality,  
 1412 resulting in single contig assemblies of the chromosomes, without gaps or  
 1413 ambiguous bases (Ns), with a total genome size comprising of 7,179,914 bp  
 1414 (including plasmids sequences, with 6,975,788 bp for the chromosome alone) for  
 1415 SM17; and 7,703,166 bp (without plasmids) for SM18 (Table 1A). High quality  
 1416 genome assemblies are highly advantageous for determining the core genome;  
 1417 identifying genome sequence and structure variants; analyzing gene acquisition  
 1418 and duplication; together with exploring the potential presence of smBGCs at a  
 1419 genetic level, which is particularly relevant for studies on the *Streptomyces* genus  
 1420 (Bentley et al., 2002; Schmid et al., 2018). Although a few marine *Streptomyces* spp.  
 1421 isolates have recently had their genomes sequenced, the majority of these consist of  
 1422 considerably fragmented sequences due to the complexity of the genome  
 1423 assemblies; which to a large extent hinders an in-depth analysis of these organisms  
 1424 at a genomic level, particularly with respect to analyzing the presence of smBGCs  
 1425 (Gomez-Escribano et al., 2015; Jackson et al., 2018). To our knowledge, this is one of  
 1426 the first studies to report the complete genome sequence of marine sponge-derived  
 1427 *Streptomyces* spp. isolates.

1428           The sequencing approach employed allowed the identification of plasmids  
 1429 in the SM17 isolate – pSM17A, pSM17B, and pSM17C (Table 1B). A series of factors  
 1430 led to their classification as plasmids, instead of simply fragments of the  
 1431 chromosome. Firstly, the contigs were much smaller than the super contig  
 1432 determined to be the chromosome: 153,923 bp, 28,056 bp, and 22,147 bp,  
 1433 respectively, when compared to 6,975,788 bp for the chromosome. In addition, their  
 1434 GC content varied from that of the chromosome, which is characteristic of  
 1435 exogenous and plasmid DNA (Nishida, 2012). The approximate sequencing  
 1436 coverage of the sequences was also varied, which is an indicator of differences in  
 1437 the copy number of the plasmid molecules, with pSM17B having a considerably

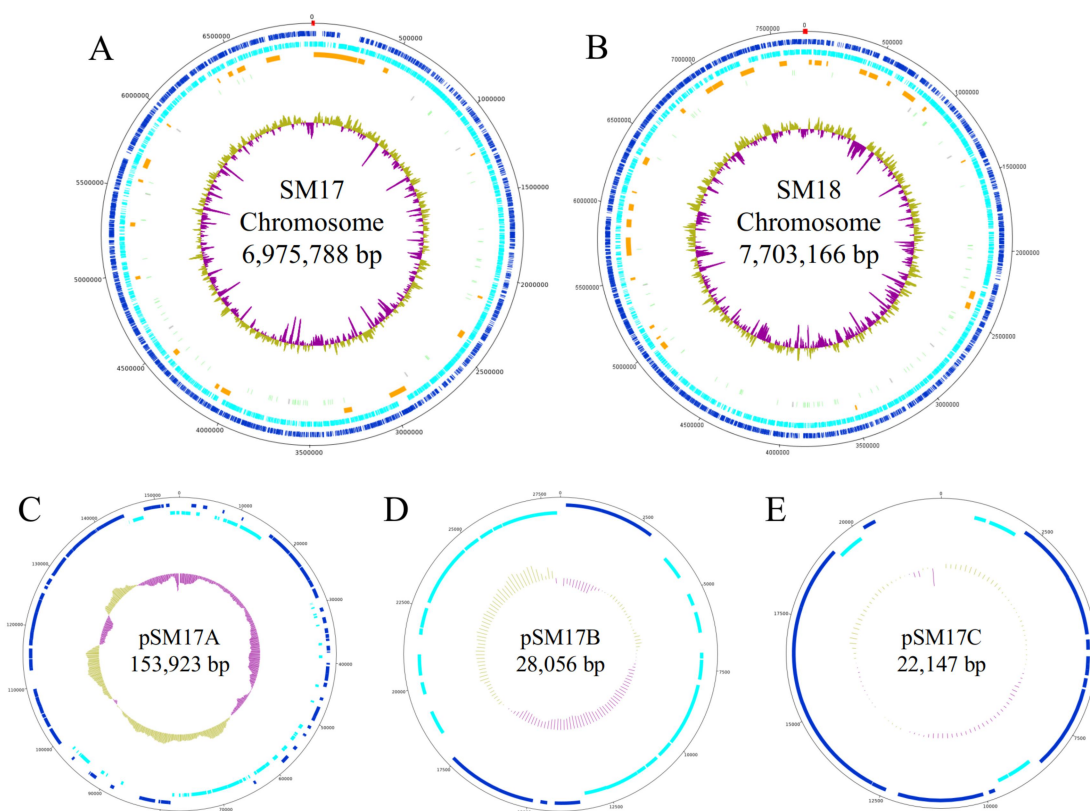
larger coverage of 548×, as opposed to 170× for pSM17A and 95× for pSM17C (Rasko et al., 2007). Finally, they were determined to share high sequence identity to other plasmids from *Streptomyces* spp. deposited in the GenBank database, as shown in Table 1B (Guo et al., 2011; Wang et al., 2012; Liu et al., 2016).

	SM17 chromosome	Plasmid pSM17A	Plasmid pSM17B	Plasmid pSM17C
Size (bp)	6,975,788	153,923	28,056	22,147
Approximate coverage (fold)	148	170	548	95
GC content (%)	73.43	69.9	72.68	74.33
Number of coding sequences (hypothetical proteins)	5,972 (2,465)	170 (145)	30 (24)	24 (21)
Top BLASTN hit	-	<i>Streptomyces</i> sp. HK1 plasmid pSHK1 (accession no. EU372836.1)	<i>Streptomyces</i> sp. Y27 plasmid pWTY27 (accession no. GU226194.2)	<i>Streptomyces</i> sp. FR-008 plasmid pSSFR2 (accession no. CP009804.1)

**Table 1B:** General characteristics of the SM17 chromosome and the three linear plasmids detected in the genome assembly

Potential Terminal Inverted Repeats (TIRs) with an estimated size of approximately 13.4 kb and 14.6 kb were identified in both the SM17 and SM18 chromosomes respectively, using a reciprocal BLASTN approach at the ends of the chromosome sequences (Gomez-Escribano et al., 2015). The *Streptomyces* genus is known to possess linear chromosomes with TIRs, with lengths varying among species; ranging from 14 bp in *Streptomyces hygroscopicus* 5008 to over 1 Mbp in *S. coelicolor* (Weaver et al., 2004; Wu et al., 2012). Although TIRs are commonly encountered in *Streptomyces* spp., their function has not yet been definitively proven, with suggested roles been proposed including chromosome stability, replication and recombination; and genome plasticity (Volff et al., 1997; Goshi et al.,

2002; Choulet et al., 2006a,b; Lin et al., 2009). The main genomic features of SM17 and the three plasmids, and SM18 (number of base pairs, coding sequences (CDSs), GC% content, and the TIRs regions) are presented in the genome maps in Figure 1.



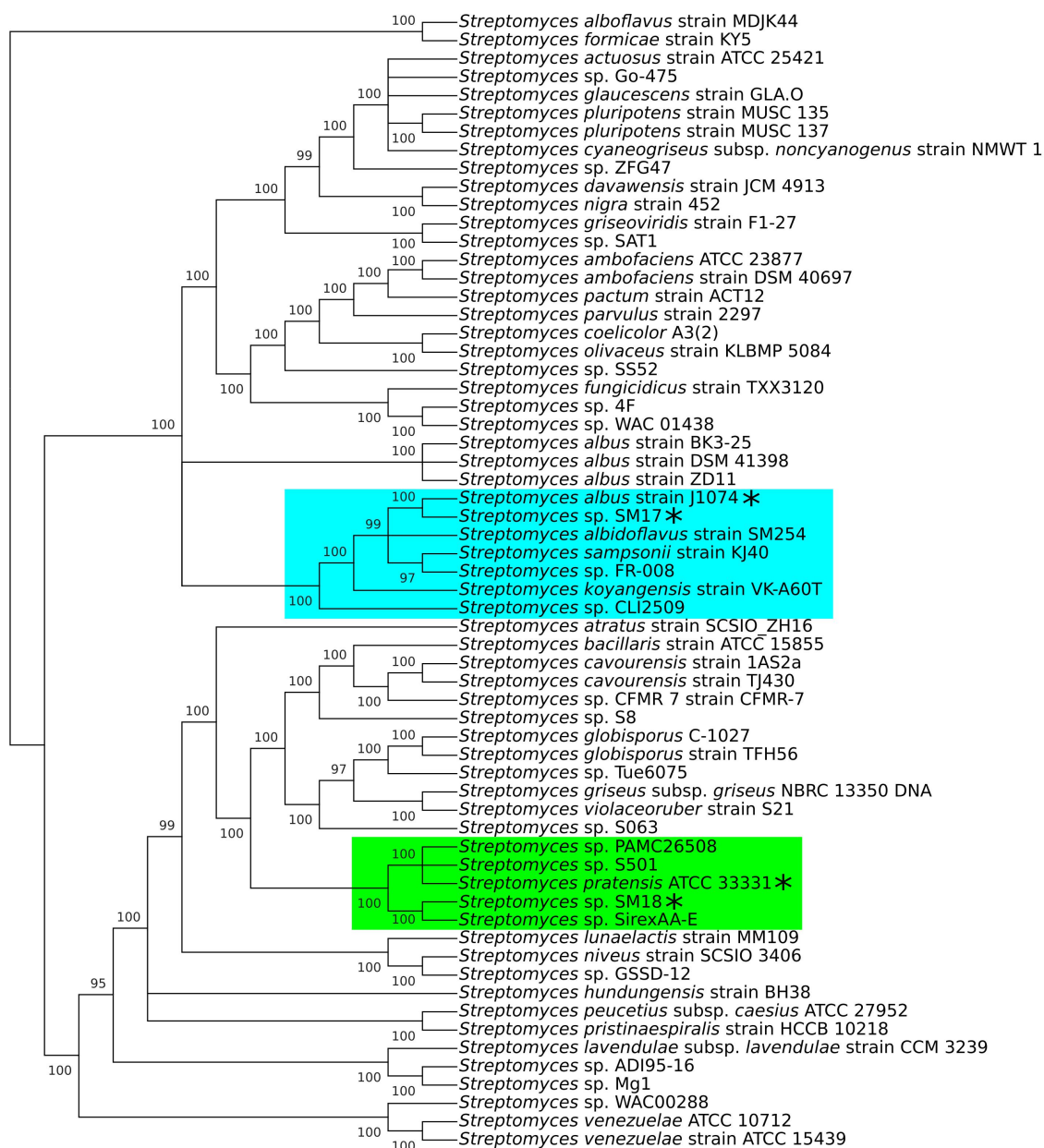
**Figure 1:** Genome maps of the SM17 and the SM18 chromosomes (A,B), and the SM17 plasmids pSM17A (C), pSM17B (D), pSM17C (E), generated using the Artemis and DNAPlotter programs. All the molecules were *in silico*-determined to be linear, although they are represented in a circular fashion, and the sizes are not representative of the scale. The following are represented from the outer to the inner circles: the nucleotide position; coding sequences (CDSs) in the forward strand (in blue); CDSs in the reverse strand (in cyan); regions of putative secondary metabolite biosynthetic gene clusters (smBGCs, in orange); tRNA and rRNA genes (in gray and green, respectively); GC% plot on default settings (above average in olive and below average in purple). In (A,B), detailed in red are the regions determined to be the terminal inverted repeats (TIRs).

## 1468 4.2. Determining the closest terrestrial type-strain relative for the marine sponge- 1469 derived isolates

1470 In order to analyze possible niche adaptations in the marine sponge-derived  
1471 SM17 and SM18 isolates, phylogenetic and whole-genome alignment analyses were  
1472 performed to identify the closest terrestrial type-strain relative, with the complete  
1473 genome sequence available in GenBank, of each isolate; with a view to performing  
1474 subsequent phenotypic, morphological and genomic comparisons once these  
1475 relatives had been determined.

1476 Phylogenetic analysis was performed using the 16S rRNA and other  
1477 housekeeping aforementioned genes, which allowed us to determine that *S. albus*  
1478 J1074 and *S. pratensis* ATCC 33331 were the closest type-strain relative to the SM17  
1479 and SM18 strains, respectively (Figure 2). Notably, SM17 and J1074 – a derivative of  
1480 the soil isolate *Streptomyces albus* G (Chater and Wilde, 1976, 1980) – are included  
1481 in the same sub-clade, while SM18 and ATCC 33331 are not, indicating that the  
1482 latter pair are more distantly related than the former. Nevertheless, further  
1483 analyses were performed with the ATCC 33331 strain, as it was the type-strain  
1484 included in the SM18 clade that was readily available in culture collections. Also, it  
1485 is important to note that the ATCC 33331 strain is the only soil-derived isolate  
1486 present in the SM18 clade (NCBI BioSample: SAMN00191232), while SirexAA-E  
1487 was isolated from an insect/microbe symbiotic community (Bianchetti et al., 2013);  
1488 PAMC26508 was isolated in association with the Antarctic lichen *Cladonia borealis*  
1489 (Shin et al., 2013); and S501 was isolated from the sediment from a seaside wetland  
1490 (NCBI BioSample: SAMN10144670). Thus, for these aforementioned reasons (being  
1491 a type-strain with its complete genome available on GenBank, isolated from soil,  
1492 and available from culture collections), the ATCC 33331 strain was determined to  
1493 be the most suitable isolate identified in the SM18 clade for the purposes of this  
1494 study.





**Figure 2:** Phylogenetic tree of the concatenated nucleotide sequence of the 16S rRNA gene, plus the housekeeping genes *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB*. Including in this analysis are the SM17 and SM18 isolates, plus 60 *Streptomyces* isolates with complete genomes available in the GenBank database. Generated using MrBayes and MEGA X, with a posterior probability cut off of 95%.

1501           To further support the similarities between our marine strains, SM17 and  
1502 SM18, and their closest terrestrial counterparts, J1074 and ATCC 33331, alignments  
1503 of the individual 16S rRNA and the other housekeeping genes were performed  
1504 with NCBI BLASTN and BLASTX (Table 2). The high identity values determined  
1505 by the analysis allowed further comparisons to be determined between the related  
1506 pairs, and also between all four *Streptomyces* strains. Notably, the identities for the  
1507 SM17- *S. albus* J1074 pair are higher (>99% for all the genes analyzed) than those  
1508 for the SM18 – *S. pratensis* ATCC 33331 pair; (91% to 99% identity depending on  
1509 the gene using BLASTN, and >95% using BLASTX). This further indicates that  
1510 SM17 and *S. albus* J1074 are very closely related organisms – possibly even  
1511 belonging to the same species, while the SM18 and *S. pratensis* ATCC 33331 are  
1512 more distantly related.

Gene	SM17-J1074	SM18-ATCC 33331
16S rRNA	1523/1524 (99%)	1519/1523 (99%)
<i>atpD</i> (ATP synthase subunit beta)	1442/1443 (99%) 480/480 (100%)	1395/1443 (97%) 454/480 (95%)
<i>gyrB</i> (DNA gyrase subunit B)	2123/2124 (99%) 706/707 (99%)	1943/2127 (91%) 683/708 (96%)
<i>recA</i> (recombinase RecA)	1123/1125 (99%) 333/333 (100%)	1058/1132 (93%) 326/332 (98%)
<i>rpoB</i> (DNA-directed RNA polymerase subunit beta)	3480/3483 (99%) 1142/1142 (100%)	3342/3487 (96%) 1116/1142 (98%)
<i>trpB</i> (tryptophan synthase beta chain)	1263/1263 (100%) 420/420 (100%)	1178/1281 (92%) 390/406 (96%)

**Table 2:** 16S rRNA and housekeeping gene alignment comparisons using the NCBI BLAST tool, between the pairs SM17 and *S. albus* J1074 (second column: SM17-J1074); and SM18 and *S. pratensis* ATCC 33331 (third column: SM18-ATCC 33331). For the housekeeping genes, the first values presented are the BLASTN (nucleotide-nucleotide) alignment identities, while the second values below are the BLASTX (translated nucleotide-protein) alignment identities. For the 16S rRNA analysis only the BLASTN alignment was performed.

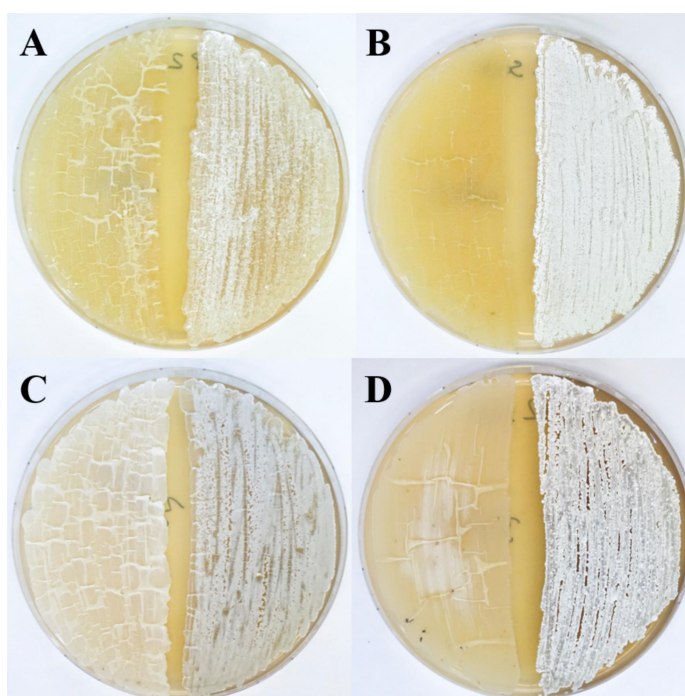
Following the 16S rRNA and housekeeping genes analyses, *S. albus* J1074 and *S. pratensis* ATCC 33331 were selected for subsequent similarity analysis using a whole-genome alignment approach with the MUMmer program (Supplementary Figure S1). Large sections of the genomes are quite well conserved between the marine sponge-derived isolates and their closest relative organism, particularly when comparing SM17 with *S. albus* J1074 (Supplementary Figure S1A). This result further confirms previous analyses, and further supports *S. albus* J1074 and *S. pratensis* ATCC 33331 as suitable terrestrial relatives, for comparative purposes.

1529 Interestingly, previous studies also reported *Streptomyces* spp. marine  
 1530 sponge-derived isolates that were determined to be closely related to *S. albus* J1074  
 1531 (Ian et al., 2014; Iniyan et al., 2016; Almeida et al., 2018). Some of these strains,  
 1532 namely PVA 94-07; GBA 94-10; and *Streptomyces albus* ICN33; were isolated from  
 1533 completely different sample types and geographic locations than those of the  
 1534 current study. While SM17, which based on the aforementioned comparative  
 1535 analysis appears to be closely related to *S. albus* J1074, was isolated from the  
 1536 sponge *Haliclona simulans* from Kilkieran Bay (Galway, Ireland), at a depth of 15  
 1537 m; the strains PVA 94-07 and GBA 94-10 were isolated from the sponges *Phakellia*  
 1538 *ventilabrum* and *Geodia barretti*, respectively; from the Tautra ridge (Trondheim  
 1539 fjord, Norway), at a depth of 121 m (Ian et al., 2014), while *Streptomyces albus*  
 1540 ICN33 was isolated from the sponge *Acanthella elongata*, from the Colachel coast  
 1541 (Kanyakumari District, Tamil Nadu), at an unspecified depth (Iniyan et al., 2016).  
 1542 This raises the possibility that “*albus*-like” *Streptomyces* strains may be  
 1543 ubiquitously associated with marine sponges.

#### 1544 4.3. Phenotype, morphology, and differential growth assessment

1545 Members of the *Streptomyces* genus are known to be capable of colonizing a  
 1546 wide variety of different ecosystems, including soil, rhizosphere, lake and marine  
 1547 sediments, and have also been reported to be associated with insects, lichen, and  
 1548 sponges (Goodfellow and Fiedler, 2010; Bianchetti et al., 2013; Rashad et al., 2015;  
 1549 Liu et al., 2017; Ay et al., 2018; Jackson et al., 2018). Thus, it is reasonable to assume  
 1550 that these organisms possess a genetic plasticity and capability that facilitates their  
 1551 adaptation to such varied environmental niches (Hoff et al., 2018). Interestingly,  
 1552 previous studies have reported that *Streptomyces* spp. isolated from marine  
 1553 environments often possess the capacity of growing independently of the presence  
 1554 of sea salts in the growth medium (Goodfellow and Fiedler, 2010; Ian et al., 2014).  
 1555 In fact, many marine isolates often display very active metabolic profiles under

1556 such conditions (Goodfellow and Fiedler, 2010). To assess whether the SM17 and  
 1557 SM18 isolates had phenotypical and/or morphological differences with respect to  
 1558 their ability to grow under different conditions, they were cultured in ISP2  
 1559 medium with and without the presence of ASW and compared with their  
 1560 terrestrial relatives (Figure 3), in a similar fashion to work previously conducted by  
 1561 the Zotchev group (Ian et al., 2014).



1562 **Figure 3:** Differential growth assessment of marine and terrestrial  
 1563 *Streptomyces* strains. From left to right, (A) *S. albus* J1074 and SM17  
 1564 on ISP2 agar medium; (B) *S. albus* J1074 and SM17 on ISP2 + ASW  
 1565 agar medium; (C) *S. pratensis* ATCC 33331 and SM18 on ISP2 agar  
 1566 medium; (D) *S. pratensis* ATCC 33331 and SM18 on ISP2 + ASW agar  
 1567 medium, following 3 days growth.

1568 All the pair-wise comparisons showed clear morphological differences  
 1569 between the marine sponge-derived isolates and their respective terrestrial  
 1570 counterparts (Figure 3). All the isolates grew effectively in the ISP2 medium

without ASW (Figures 3A,C), even though there were slight differences regarding growth and sporulation; with the SM17 isolate being able to grow and sporulate more rapidly in comparison to *S. albus* J1074 (Figure 3A). There was no clear difference in the growth of SM18 and *S. pratensis* ATCC 33331 on the ISP2 growth medium without ASW, although they clearly displayed very different morphological features (Figure 3C). On the other hand, when grown on the ISP2 medium with ASW, *S. albus* J1074 was clearly less capable of growing in the presence of sea salts, while SM17 thrived (Figure 3B). This result is particularly interesting, since, as previously shown (Figure 2, Table 2, and Supplementary Figure S1), these two organisms are genetically very similar. In contrast, there were less marked differences in the ability of both SM18 and *S. pratensis* ATCC 33331 to grow in the presence of sea salts (Figure 3D). While SM18 appeared to grow better, nevertheless *S. pratensis* ATCC 33331 was still able to grow in the ISP2 medium containing sea salts albeit more slowly than SM18; and indeed, more slowly than when *S. pratensis* ATCC 33331 was cultured in the absence of ASW (Figure 3C). From these observations, it became clear that a more thorough analysis of the SM17 and SM18 genomes might provide some interesting insights regarding potential genome-wide adaptations that may have occurred in these marine isolates, which may have resulted in them being able to grow more efficiently in the ISP2 medium supplemented with ASW; relative to their terrestrial counterparts.

#### 4.4. Prediction of secondary metabolite biosynthetic gene clusters (smBGCs)

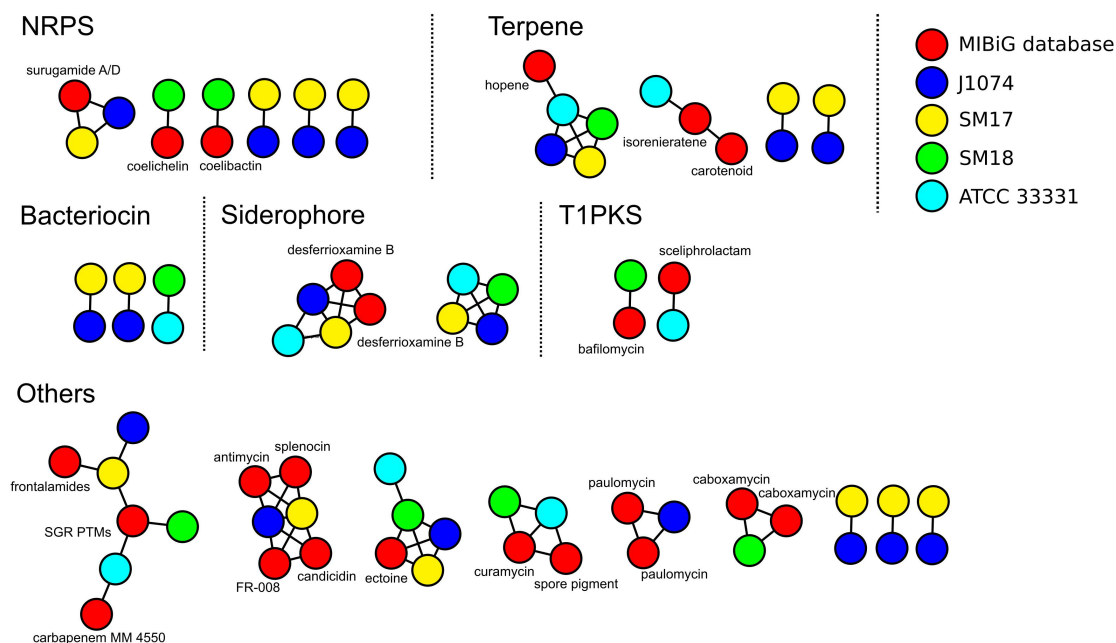
Members of the Actinomycetales order are historically known to produce a broad range of bioactive compounds of biotechnological and clinical interest, and among them, the *Streptomyces* genus excels, with over 10,000 bioactive compounds produced by members of the genus being discovered to date (Hwang et al., 2014; Ziemert et al., 2016; Kamjam et al., 2017; Lee et al., 2018). The marine sponge-derived SM17 and SM18 strains have previously been reported to possess

antimicrobial activity against gram-negative and gram-positive bacteria – including the methicillin-resistant *S. aureus* (MRSA), and yeasts (Kennedy et al., 2009; Jackson et al., 2018). To provide insights at a genomic level regarding which compounds might be responsible for the previously observed antimicrobial activity, we employed the antiSMASH program in an attempt to predict the presence of putative smBGCs, based on homology to known smBGCs deposited in the databases (Blin et al., 2017). Several gene clusters were predicted to be present in both SM17 and SM18 (Supplementary Tables S1, S2), with a total of 20 potential smBGCs in SM17, and 26 in SM18; with a variety of cluster types being assigned, including: type I polyketide synthases (T1pks), type II polyketide synthases (T2pks), type III polyketide synthases (T3pks), non-ribosomal peptide synthetases (NRPS), lantipeptides, bacteriocins, and terpenes. These types of clusters are known to produce a variety of compounds with antimicrobial activity, including: erythromycin (T1pks); tetracenomycin (T2pks); germicidin (T3pks); daptomycin (NRPS); nisin (lantipeptide/lantibiotic/bacteriocin); and pentalenolactone (terpene) (Shen, 2003; Tetzlaff et al., 2006; Robbel and Marahiel, 2010; Shi et al., 2011; Yamada et al., 2015; Čihák et al., 2017).

The antiSMASH predictions were also further analyzed using the BiG-SCAPE program (Navarro-Muñoz et al., 2018), which allowed us to cluster the predicted smBGCs into gene cluster families (GCFs) based on their sequences and Pfam protein families similarities (El-Gebali et al., 2019), and also to compare them to known smBGCs available from the latest version of the MIBiG repository (version 1.4) (Medema et al., 2015), which can also assist in improving the annotations of the predicted smBGCs. Based on their similarity to known smBGCs, some of the bioactive compounds predicted to be encoded by these smBGCs may be compatible with the previously determined antimicrobial capabilities of the SM17 and SM18 isolates (Figure 4 and Supplementary Tables S1, S2). For example,

SM17 appears to possess a candicidin, an antimycin, and a polycyclic tetramate macrolactam cluster (SGR PTMs) (Figure 4 and Supplementary Table S1), with similarity to the candicidin, antimycin and tetramate macrolactam sequences from *Streptomyces* sp. FR-008, *Streptomyces* sp. S4 and *Streptomyces griseus* in the database, respectively, and which are known to have anti-fungal properties (Campelo and Gil, 2002; Chen et al., 2003; Seipke et al., 2011; Luo et al., 2013). SM17 also contains clusters that may potentially encode for the production of surugamides (Figure 4) and the glycopeptide antibiotic mannopeptimycin (Supplementary Table S1), with the former possessing gene similarity with the surugamide A/D sequence from *Streptomyces albus* in the database (Ninomiya et al., 2016), and the latter sharing similarity to the mannopeptimycin sequence from *Streptomyces hygroscopicus* in the database, with the main biosynthetic genes being present in the predicted smBGC (Singh et al., 2003; Magarvey et al., 2006). SM18 appears to possess a cluster encoding the anti-bacterial compound bafilomycin (Figure 4 and Supplementary Table S2), with similarity to the bafilomycin sequence from *Streptomyces lohii* (Bowman et al., 1988; Zhang et al., 2013; Nara et al., 2017); as well other clusters with similarity to known smBGCs that encode anti-fungal and anti-bacterial compounds such as SGR PTMs, curamycin, and caboxamycin (Figure 4), from *Streptomyces griseus* (Luo et al., 2013), *Streptomyces curacoii* (Galmarini and Deulofeu, 1961), and *Streptomyces* sp. NTK 937 (Hohmann et al., 2009; Losada et al., 2017), respectively.

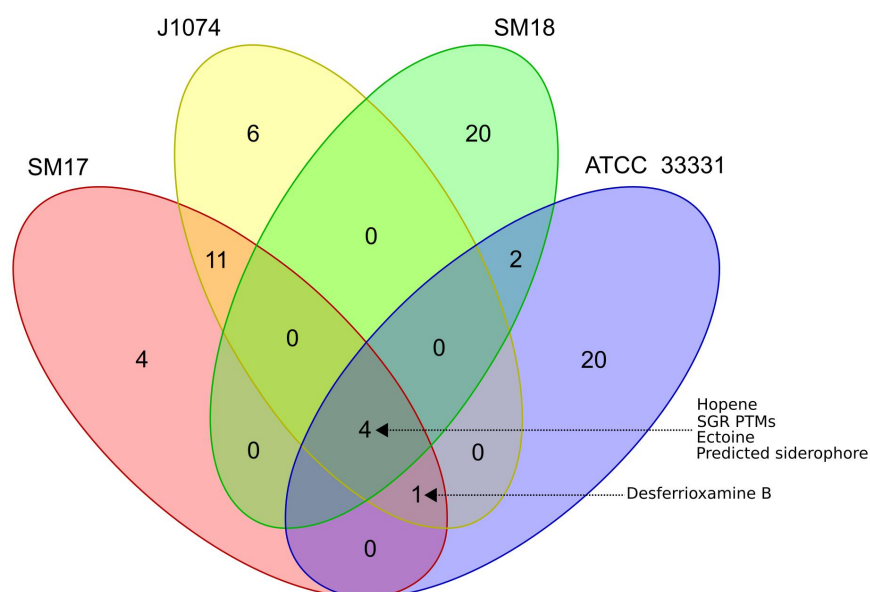




**Figure 4:** Gene clusters families (GCFs) analysis using antiSMASH (version 4), BiG-SCAPE (version 20181005), MIBiG database (version 1.4), and Cytoscape. Each node represents a smBGC predicted in the respective organism (labeled in different colors), and the interactions represent cluster similarity. Annotations of the MIBiG database smBGCs are labeled accordingly. Singletons, i.e., smBGCs without similarities with the smBGCs in the MIBiG database, or without similarities with the smBGCs predicted in the other genomes analyzed in this study, are not included in this figure.

We also performed the antiSMASH and BiG-SCAPE analysis on *S. albus* J1074 and *S. pratensis* ATCC 33331 genomes, in an effort to determine to what extent the marine sponge-derived isolates SM17 and SM18 may potentially produce similar and/or unique compounds when compared to their terrestrial counterparts (Figure 4 and Supplementary Tables S3, S4). Based on the BiG-SCAPE similarity clustering, a Venn diagram was generated, representing the presence/absence of GCFs in the SM17, SM18, *S. albus* J1074, and *S. pratensis* ATCC

33331 genomes (Figure 5). In keeping with the phylogeny results which indicated that SM17 and *S. albus* J1074 were very closely related organisms, the smBGCs predictions and similarity clustering results were also strikingly similar (Figures 4, 5). Among a total of 42 predicted smBGCs in both genomes (22 in *S. albus* J1074 and 20 in SM17), 10 seem to be unique (6 in *S. albus* J1074, and 4 in SM17) (Figure 5). In contrast, there was a much larger number of predicted unique smBGCs between SM18 and *S. pratensis* ATCC 33331, where amongst a total of 53 predicted clusters (27 in *S. pratensis* ATCC 33331 and 26 in SM18), only 6 appear to be present in both genomes; with the majority being potentially unique (20 in *S. pratensis* ATCC 33331 and 20 in SM18) (Figure 5). Also, a total of 4 smBGCs were shared among all of the strains analyzed (Figure 5), and these were determined to be: hopene; SGR PTMs family of smBGCs, ectoine, and a predicted siderophore smBGC without significant similarity to sequences in the MIBiG database (Figures 4, 5).



**Figure 5:** Venn diagram representation of GCFs presence/absence analysis using BiG-SCAPE.

1678           Notably, smBGCs encoding the production of desferrioxamines, which are  
1679 hydroxamate siderophores, while present in *S. albus* J1074, *S. pratensis* ATCC  
1680 33331 and SM17 (Figures 4, 5), are absent in the SM18 genome (Supplementary  
1681 Table S2 and Figure 4). Siderophores are specialized metabolites that function to  
1682 scavenge  $\text{Fe}^{3+}$ , and hence are crucial for sessile organisms to assimilate iron (Hider  
1683 and Kong, 2010). Genes involved in desferrioxamines production, in particular, are  
1684 widely conserved in marine microorganisms, and are believed to be present in all  
1685 *Streptomyces* species (Tierrafría et al., 2011; Cruz-Morales et al., 2017). Thus, this  
1686 may be the first report of a *Streptomyces* isolate that does not possess a smBGC  
1687 that encodes for the production of desferrioxamines. The SM18 isolate does,  
1688 however, possess smBGCs encoding other siderophores, such as coelichelin, and  
1689 mirubactin (Supplementary Table S2 and Figure 4), which may circumvent for the  
1690 lack of production of desferrioxamines with respect to iron acquisition in the strain.

1691           During processing of the data for this study, a newer version of the  
1692 antiSMASH webserver (version 5) was released (Blin et al., 2019). Using this new  
1693 version of antiSMASH did not result in any major differences being detected in the  
1694 data being analyzed, it did however result in the identification of a smBGC  
1695 encoding mycemycin in the SM18 genome. Mycemycin is a relatively newly  
1696 identified compound, from marine and soil *Streptomyces* isolates, belonging to the  
1697 dibenzoxazepinone (DBP) family, which possesses HIV-1 reverse transcriptase  
1698 inhibitory activity (Liu et al., 2015; Song et al., 2018). Production of the DBP family  
1699 of compounds appear to date to be rare in the microbial world, and these  
1700 compounds possess a broad range of interesting activities, including anti-HIV and  
1701 anti-tumor activities (Zhang et al., 2018). Thus, pursuing the identification of new  
1702 members of this family of compounds may be worthwhile, and it is interesting to  
1703 report the potential presence of a smBGC encoding the production of mycemycin  
1704 in another *Streptomyces* isolate.

Nevertheless, it is clear that further analysis would need to be undertaken to confirm that these compounds are in fact being produced by the SM17 and SM18 isolates, as some of these smBGCs are likely to be cryptic and the compounds may not be produced under certain culture conditions (Rutledge and Challis, 2015; Rigali et al., 2018). Given that SM17 and *S. albus* J1074 are genetically very similar, it is perhaps reasonable to expect that regulation of secondary metabolite production may to some extent be similar in both strains. Therefore, it may be possible to use what is currently known about the better studied *S. albus* J1074 isolate to gain a better understanding regarding the expression of certain smBGCs and the metabolic pathways involved in SM17 (Hoz et al., 2017; Kallifidas et al., 2018; Nguyen et al., 2018).

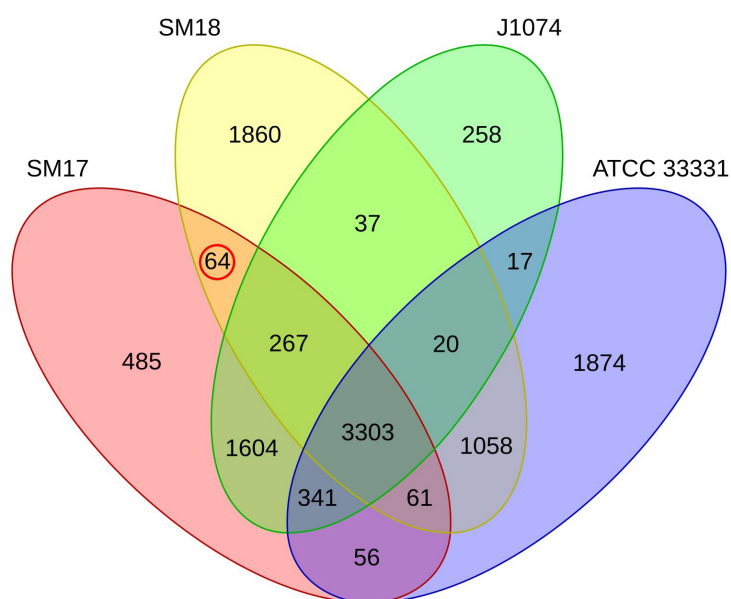
#### 4.5. Comparative genomics

A series of comparative genomics analyses was then performed in order to further characterize the marine sponge-derived isolates SM17 and SM18 at the genome level, and in particular to compare them to their respective closest terrestrial relative.

##### 4.5.1. Analysis of orthologous genes

The Roary program was used to determine the pan-genome; the core genome; the accessory genome; and the strain-specific genome (the genes that are uniquely present in only one of the isolates), in the marine sponge-derived isolates *Streptomyces* sp. strain SM17 and *Streptomyces* sp. strain SM18, and their respective closest terrestrial relatives *S. albus* J1074 and *S. pratensis* ATCCC 33331 (Figure 6) (Page et al., 2015). The pan-genome was determined to consist of 11,305 genes; while the core genome consisted of 3,303 genes (~29% of the pan-genome); and the accessory genome consisted of 8,002 genes (~71% of the pan-genome). For the strain-specific genomes, SM17 had 485 unique genes; SM18 had 1,860 unique

1731 genes; *S. albus* J1074 had 258 unique genes; and *S. pratensis* ATCC 33331 had 1,874  
1732 unique genes. This is a combined total of 4,477 unique genes (~39% of the pan-  
1733 genome, and ~56% of the accessory genome). Notably, in keeping with what we  
1734 had previously observed with the phylogeny and whole-genome alignment  
1735 analyses, the SM17 and J1074 strains shared a very large number of orthologous  
1736 genes (a total of 5,515 shared genes, or ~89% and ~94% of the SM17 and the J1074  
1737 total number of CDSs, respectively), further indicating that they are very closely  
1738 related organisms. In contrast, SM18 and *S. pratensis* ATCC 33331 shared a much  
1739 lower proportion of their genes: 4,469 genes (or ~67% and ~66% for the SM18 and  
1740 ATCC 33331 total number of CDSs, respectively). A total of 64 orthologous genes  
1741 were found to be commonly present in the marine sponge-derived isolates SM17  
1742 and SM18, while absent in their terrestrial counterparts J1074 and ATCC 33331  
1743 (Figure 6). Given that they are absent in both terrestrial relatives, we undertook  
1744 further analyses of these genes to assess their potential function(s) in an effort to  
1745 provide insights into potential ENAs in both these sponge-derived isolates.



**Figure 6:** Venn diagram representing the presence/absence of orthologous genes in the SM17, SM18, *S. albus* J1074, and the *S. pratensis* ATCC 33331 genomes. Orthologous genes that are present commonly in the marine sponge-derived isolates SM17 and SM18, while absent in their terrestrial counterparts J1074 and ATCC 33331, are circled in red.

#### 4.5.2. Orthology analysis of smBGC-associated genes

The genes previously determined to be associated with smBGCs, using the antiSMASH program, were subsequently analyzed using the Roary program, to identify smBGCs-associated genes which were shared or unique between the four genomes (Supplementary Figure S2) (Page et al., 2015; Blin et al., 2017). With respect to potential smBGCs-associated genes, very few genes appeared to be conserved in all the organisms (a total of 58 genes, corresponding to 2.8% of the total smBGCs-associated gene pool, or 0.017% of the core genome) (Supplementary Figure S2). The largest number of unique smBGCs-associated genes is present in

1761 the SM18 isolate (623 genes), followed by *S. pratensis* ATCC 33331 (485 genes)  
1762 which may be indicative of a greater potential to produce diverse secondary  
1763 metabolites in these isolates. In contrast, SM17 and *S. albus* J1074 appear to possess  
1764 a lower quantity of unique smBGCs-associated genes; with 132 and 150 unique  
1765 genes, respectively (Supplementary Figure S2).

1766 Interestingly, when comparing the Venn diagrams from Figure 6 and  
1767 Supplementary Figure S2, it appears that a large portion of the unique genes  
1768 present in the isolates are potentially related to the production of secondary  
1769 metabolites. For SM17, ~27% (132 out of 485) of the unique genes are potentially  
1770 smBGCs-associated genes, while for SM18 this percentage is ~33% (623 out of 1860);  
1771 ~58% (150 out of 258) for *S. albus* J1074; and ~26% (485 out of 1874) for *S. pratensis*  
1772 ATCC 33331. Taken together, these results indicate that, even for closely related  
1773 *Streptomyces* spp. isolates (particularly when considering the pair SM17 and *S.*  
1774 *albus* J1074), there is still potential to discover different secondary metabolites from  
1775 these strains, with potentially unique characteristics. Previous reports have also  
1776 indicated that the use of closely related *Streptomyces* strains to identify new  
1777 smBGCs is useful for the identification of novel specialized biosynthetic pathways  
1778 (Antony-Babu et al., 2017; Vicente et al., 2018).

#### 1779 **4.5.3. Groups of orthologous genes commonly present in the marine sponge-** 1780 **derived isolates**

1781 Given that the SM17 and SM18 were isolated from a marine sponge and  
1782 have been shown to be more adapted to higher salinity medium (Figure 3), it is  
1783 likely that the identification of genes that are commonly present in SM17 and SM18  
1784 but not in their terrestrial relatives J1074 and ATCC 33331 may help in the  
1785 identification of potential ENAs that these strains might possess, at a genetic level.  
1786 The previous analysis of orthologous genes allowed us to determine which groups  
1787 of orthologous genes are present commonly in the marine sponge-derived isolates

1788 SM17 and SM18, while absent in their terrestrial counterparts J1074 and ATCC  
1789 33331, as highlighted in Figure 6. This was performed by taking into account  
1790 sequence homology and gene synteny (e.g., splitting paralogous genes with the  
1791 Roary program); hence different copies of a gene can belong to different  
1792 orthologous group due to potentially different evolutionary events such as gene  
1793 duplication or lateral gene transfer occurring in the *Streptomyces* genomes (Zhou  
1794 et al., 2012; Page et al., 2015). Thus, from here on the orthologous genes will be  
1795 referred to simply as “genes.” In doing this we identified a potential ENA gene  
1796 pool which consisted of 64 genes (Table 3). These were then manually annotated  
1797 using the NCBI GenBank, CDD, UniProt, and the InterPro databases (Johnson et al.,  
1798 2008; Marchler-Bauer et al., 2011; UniProt Consortium, 2015; Benson et al., 2018;  
1799 Mitchell et al., 2018), and hypothetical proteins were removed, resulting in a final  
1800 total of 57 genes (Supplementary Table S5). The ENA gene pool included functional  
1801 categories of genes that are likely to be related to niche adaptations in the marine  
1802 sponge-derived isolates, and included a total of 29 genes that could be grouped  
1803 based on potential biological functions such as osmotic stress defense;  
1804 transcriptional regulation; symbiotic interactions; antimicrobial compounds  
1805 production and resistance; ABC transporters; together with horizontal gene  
1806 transfer and defense-related features (Table 3).



Environmental niche adaptation	Gene name	Product
Osmotic stress defence	<i>nuoA</i>	NADH-quinone oxidoreductase subunit A
	<i>nuoH</i>	NADH-quinone oxidoreductase subunit H
	<i>nuoJ</i>	NADH-quinone oxidoreductase subunit J
	<i>nuoK</i>	NADH-quinone oxidoreductase subunit K
	<i>nuoL</i>	NADH-quinone oxidoreductase subunit L
	<i>nuoM</i>	NADH-quinone oxidoreductase subunit M
	<i>nuoN</i>	NADH-quinone oxidoreductase subunit N
	<i>proP</i>	Proline/betaine transporter
Transcriptional regulation	<i>bepR*</i>	HTH-type transcriptional repressor BepR / TetR family transcriptional regulator
	<i>cynR</i>	HTH-type transcriptional regulator CynR / LysR family transcriptional regulator
	<i>degU</i>	Transcriptional regulatory protein DegU / DNA-binding response regulator
	group_5796	Transcriptional regulator, IclR family
	group_5819	Transcriptional regulator PadR-like family protein
	<i>rhmr</i>	HTH-type transcriptional regulator KipR / MarR family transcriptional regulator
	<i>tcrA</i>	Transcriptional regulatory protein CutR / DNA-binding response regulator
Symbiotic interactions	group_5772	Tetratricopeptide repeat protein
Antimicrobial compounds production and resistance	<i>aprX*</i>	Serine protease AprX / Subtilase family protein / Peptidase S8
	group_5198	Aminoglycoside phosphotransferase
	group_5385	Aminoglycoside phosphotransferase

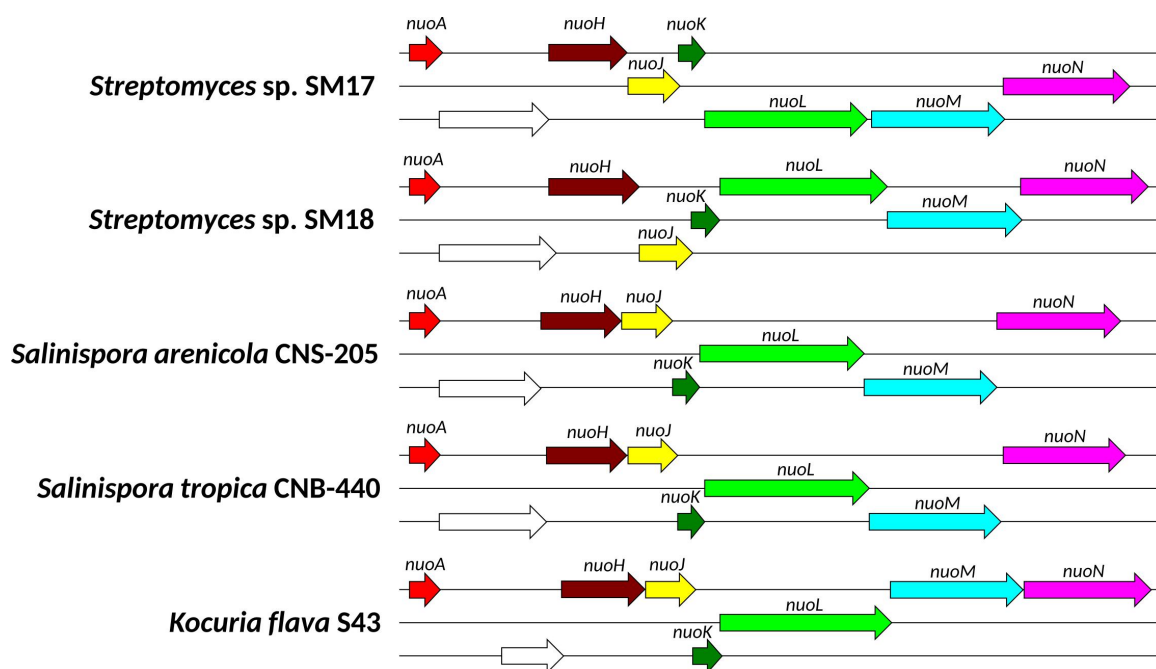
	group_5818	Acyltransferase 3
	group_5836	Acyltransferase
	<i>liaS</i>	HPK7 family sensor histidine kinase LiaS
ABC transporters	group_5821	ABC transporter permease
	<i>tauB</i>	Aliphatic sulfonates import ATP-binding protein SsuB / ABC transporter ATP-binding protein
	<i>yknY</i>	Uncharacterised ABC transporter ATP-binding protein YknY
Horizontal gene transfer and defence-related features	group_1044	Integrase core domain / IS3 family transposase
	group_1272	Toxin-antitoxin system, RelE family
	group_1944	Restriction endonuclease
	group_1945	IS3 family transposase

**Table 3:** Groups of orthologous genes and their respective annotations (excluding hypothetical proteins), which are present commonly in the sponge-derived isolates SM17 and SM18, while absent in their terrestrial counterparts J1074 and ATCC 33331. When the gene name was not determined, a generic unique name was given (e.g., group\_5796) by the Roary program. \*genes without multiple copies or paralogs in the terrestrial isolates' genomes, taking into consideration only those with a defined gene name.

### 1815 4.5.3.1. Resistance to osmotic stress

1816 For bacteria to survive in marine environments where salinity levels of  
 1817 approximately 3.5% exist, they must be able to simultaneously overcome stresses  
 1818 due to both high osmotic pressure and high Na<sup>+</sup> concentrations (Yaakop et al.,  
 1819 2016); together with other stresses including pressure, temperature and  
 1820 oligotrophic conditions (Xie et al., 2018). Bacteria typically respond to variations in  
 1821 external osmotic pressure by accumulating or releasing solutes, thereby  
 1822 attenuating water fluxes and maintaining cellular homeostasis (Wood, 2015). The  
 1823 marine sponge-derived isolates SM17 and SM18 appeared to grow and  
 1824 differentiate more rapidly when grown on media containing artificial seawater,  
 1825 when compared to their closely related terrestrial counterparts (Figure 3), thus  
 1826 indicating a potential increased fitness to higher salinity environments, as also  
 1827 previously described in other marine *Streptomyces* isolates (Ian et al., 2014).  
 1828 Previous studies with marine Actinomycetes, specifically with the genera  
 1829 *Salinispora*, *Streptomyces*, and *Kocuria*, have proposed that the NADH-quinone  
 1830 oxidoreductases *nuoAHJKLMN* genes, which encode a proton pump, could be  
 1831 classified as potential MAGs (Penn and Jensen, 2012; Ian et al., 2014; Sun et al.,  
 1832 2018). This proton pump is believed to create a proton-motive force which  
 1833 generates ATP, helping to maintain a proton gradient in seawater (Penn and Jensen,  
 1834 2012; Ian et al., 2014; Sun et al., 2018). We identified the *nuoAHJKLMN* genes in the  
 1835 ENA gene pool in both SM17 and SM18 (Table 3). Further analysis indicated that  
 1836 both isolates possessed one extra copy of these genes when compared to their  
 1837 terrestrial counterparts, and that these genes were organized in an operon-like  
 1838 structure, similar to that previously reported in *Salinispora arenicola* CNS-205 and  
 1839 in *Kocuria flava* S43 (Sun et al., 2018). Furthermore, the same gene synteny for the  
 1840 partial *nuo*-operon was present in *Streptomyces* sp. SM17, *Streptomyces* sp. SM18,  
 1841 *Salinispora arenicola* CNS-205, *Salinispora tropica* CNB-440, and *Kocuria flava* S43

(Figure 7); with *nuoA*, followed by a hypothetical protein, and then followed by *nuoH*, *nuoJ*, *nuoK*, *nuoL*, *nuoM*, and *nuoN*. It is important, however, to note that differences in sequence identity and reading frames are present (Supplementary Table S6 and Figure 7), which may indicate that different evolutionary events may have occurred in the aforementioned genomes. The presence of this partial *nuo*-operon in the sponge derived SM17 and SM18 isolates and in the other marine actinomycetes (*Salinispora arenicola* CNS-205, *Salinispora tropica* CNB-440, and *Kocuria flava* S43), which are absent in their terrestrial counterparts J1074 and ATCC 33331, may explain, at least in part, the increased tolerance to salinity we observed in SM17 and SM18 relative to J1074 and ATCC 33331; which although still able to grow in the presence of ASW, grew much more slowly (Figure 3). Another important mechanism which bacteria employ as a defense mechanism against osmotic stress is both the synthesis and the uptake of compatible solutes, such as proline, glycine, betaine and ectoine, in order to maintain membrane turgor pressure (Krämer, 2010; Lim and Lee, 2015). Extra copies of the *proP* gene, which encodes a potential proline/betaine transporter (ProP), were found in both the SM17 and SM18 strains. It has been shown in *E. coli* that the ProP transporter acts both as an osmoregulator and as an osmosensor; and is capable of transporting proline, glycine betaine, proline betaine, carnitine, ectoine and other compounds (MacMillan et al., 1999; Roessler and Muller, 2001; Burg and Ferraris, 2008). Therefore, the *proP* genes may also be related to the increased capacity of the SM17 and SM18 strains to tolerate hyperosmotic environments, as evidenced by their growth on the ASW medium (Figure 3).



**Figure 7:** Graphical representation of the gene synteny of the partial *nuo*-operon present in the genomes of the marine isolates *Streptomyces* sp. SM17, *Streptomyces* sp. SM18, *Salinispora arenicola* CNS-205, *Salinispora tropica* CNB-440, and *Kocuria flava* S43, while absent in the terrestrial isolates *Streptomyces albus* J1074 and *Streptomyces pratensis* ATCC 33331. Each of the three lines represent a reading frame and the arrows represent a gene, with their respective gene names. Genes with the same color are homologs, while the ones in white are hypothetical proteins with no homologs in the UniProt or PDB databases.

#### 4.5.3.2. Antimicrobial compounds production and resistance

For many years the main ecological function of antibiotics production in bacteria in natural environments was believed to be inhibition of the growth of other microorganisms, thereby conferring a selective advantage on the producing strain with respect to colonization of particular environmental niches (Linares et al.,

2006). In this respect, antibiotic production may be employed as a defense mechanism for the *Streptomyces* spp. isolates SM17 and SM18 – and other members of the symbiotic community – against other competitor microorganisms in the marine sponge host; as has been previously reported to be the case with other antibiotic producing microorganisms, such as *Streptomyces* spp. which have been isolated from different hosts including plants and insects (Bondarev et al., 2013; van der Meij et al., 2017; Ceapă et al., 2018; Engl et al., 2018). Furthermore, antibiotics may also play an important role in the overall defense of the sponge host itself by protecting it against pathogens, in a biological interaction defined as defensive symbiosis (Clay, 2014), which has been reported in a number of systems, including beewolf wasps and antibiotic-producing *Streptomyces* bacteria (Engl et al., 2018). Nevertheless, more recently it has been proposed that in natural environments antibiotics may also act as small molecules with signaling functions, functioning in a similar fashion to quorum sensing molecules; acting for example to alter the expression of genes; to induce biofilm formation; or to modulate colony morphology – all of which may be important in coordinated communication within symbiotic communities (Romero et al., 2011). Thus, antibiotics may have a number of roles in a niche environment such as in marine sponges, which may include both defense-related and signaling roles (Linares et al., 2006; Romero et al., 2011). The presence of a wide variety of predicted smBGCs in both SM17 and SM18 – many of which are potentially involved in the production of antimicrobial compounds (Supplementary Tables S1, S2), coupled with the previously reported antimicrobial activities in these strains (Kennedy et al., 2009; Jackson et al., 2018); supports a possible role for these two *Streptomyces* spp. isolates in defensive symbiosis in *Haliclona simulans*, from which they were isolated. In this respect two acyltransferase genes potentially involved in the biosynthesis of type II PKS antibiotics, or type I PKSs that require discrete acyltransferase enzymes, were present in the ENA gene pool (Table 3) (Cheng et al., 2003; Zhang et al., 2017). In

1908 addition, a subtilase-like serine protease gene (*aprX*) was also identified in the ENA  
 1909 gene pool (Table 3), which belongs to a family of proteins that are known to play a  
 1910 number of different biological roles, including involvement in the biosynthesis of  
 1911 antimicrobial peptides, with some possessing algicidal properties, which could  
 1912 potentially be relevant from a sponge defense perspective (Lee et al., 2000; Barra et  
 1913 al., 2017; Montalbán-López et al., 2018). Protease producing marine bacteria are  
 1914 known to be important in the degradation of organic nitrogen which is essential for  
 1915 nitrogen recycling in marine sediments (Zhang et al., 2015). Marine bacterial  
 1916 proteases are also known to play a role in sponge host nitrogen metabolism, which  
 1917 may also explain the presence of this specific protease in the marine isolates SM17  
 1918 and SM18, and its absence in their terrestrial counterparts (Li et al., 2016; Kiran et  
 1919 al., 2018). Further work on these specific proteases might also be relevant from an  
 1920 industrial perspective, given the interest in proteases of marine origin which are  
 1921 typically cold adapted, salt tolerant, with broad optimal pH values (Li et al., 2016)  
 1922 and which are particularly suited for a number of biotechnological applications,  
 1923 including laundry detergents, food processing, the leather and textile industries,  
 1924 and in waste water treatment applications (Li et al., 2013; Salwan and Sharma,  
 1925 2018).

1926 A LiaS-encoding gene was also present in the ENA pool, which has  
 1927 previously been reported to be part of the two-component LiaS/LiaR regulatory  
 1928 system, a stress-sensing module that is conserved in *Firmicutes* bacteria and which  
 1929 is involved in the response to a subset of cell wall-active antibiotics such as  
 1930 bacitracin and vancomycin in *Bacillus subtilis*; while also being involved in  
 1931 response to cationic antimicrobial peptides and secretion stress (Mascher et al.,  
 1932 2004). In *Listeria monocytogenes*, the LiaS/R system also plays an important role in  
 1933 resistance to the food preservative nisin (Collins et al., 2012). The presence of  
 1934 antibiotic resistance-related genes in our two sponge-derived isolates may be

significant from two perspectives. Firstly, they may function as part of a self-resistance mechanism in these strains, allowing them to be protected from the antimicrobial compounds that they themselves are producing; and/or secondly, as a resistance mechanism to protect themselves from the antimicrobial compounds produced by other microorganisms within the sponge symbiotic community (Wright, 2005, 2012).

#### 4.5.3.3. ABC transporters

ATP-binding cassette (ABC) transporters are ATP-dependent protein complexes that are widespread in all forms of life and which are vital in mediating the transport of both organic and inorganic molecules across cell membranes (ter Beek et al., 2014; Wilkens, 2015). In bacteria, they confer resistance to antibiotics and to other toxic compounds through efflux/transport mechanisms (Greene et al., 2018); and are also involved in nutrient acquisition and in helping to maintain osmotic balance in the cell (Wood, 2007; Fan et al., 2013; Teichmann et al., 2018). The ENA gene pool includes an *yknY*-like ABC transporter (Table 3), which has been reported to be involved in the efflux of the sporulation-delaying protein (SDP) in *Bacillus* spp., although it is still poorly characterized in other genera, such as in *Streptomyces* (González-Pastor et al., 2003; Xu et al., 2016; Greene et al., 2018). The SDP protein is a killing factor exported by cells that have started the sporulation process, therefore inducing the lysis of sister cells, making more nutrients available, and ultimately delaying the sporulation process and maintaining regular cell growth (González-Pastor et al., 2003). Thus, it is reasonable to assume that the bacterial members of the sponge symbiotic community may employ similar mechanisms and resistance genes targeting these potentially harmful proteins, which may be the case in both SM17 and SM18. A *tauB/ssuB*-like ABC transporter was also present in the ENA gene pool, which may be responsible in allowing more



versatile nutrient acquisition and cycling – specifically for nitrate and sulfonate – for the marine *Streptomyces* isolates SM17 and SM18, as it has been previously suggested to be the case for marine sponge symbiotic communities, through metagenome binning analysis (Karimi et al., 2018).

#### 4.5.3.4. Transcriptional regulation

Being able to efficiently respond to changes in their environment is crucial in helping bacteria adapt to and survive within these environments (Feklístov et al., 2014; Daniel-Ivad et al., 2018); and, as previously mentioned, it is particularly important for the sponge-derived bacteria to be able to react appropriately to osmotic and other environmental stresses such as the presence of antibiotics and other potentially harmful compounds; the lack of nutrients; or allowing cell-to-cell communication through quorum sensing.

Transcriptional regulators play a crucial role in allowing bacteria to respond appropriately to numerous environmental stimuli and are believed to be intrinsically linked to lifestyle and environmental adaptation in bacteria (Stock et al., 1990; Feklístov et al., 2014; Daniel-Ivad et al., 2018). Since the SM17 and SM18 isolates inhabit the same niche environment and are subsequently exposed to similar conditions, it is likely that they employ similar adaptive mechanisms in response to those conditions. The ENA gene pool does include a range of transcriptional regulators (Table 3), further indicating that the marine sponge-derived isolates SM17 and SM18 share signal transduction mechanisms that are absent in their terrestrial counterparts, which may account for important niche adaptations that have been acquired. Notably, the TetR, LysR, DegU, IclR, PadR, and CutR families of transcriptional regulators are present in the ENA gene pool. These are commonly associated with mechanisms that could also be potential adaptations employed by the sponge-derived isolates SM17 and SM18, such as for

example: antibiotics production (TetR, DegU); antibiotics resistance (TetR, LysR); multidrug resistance (IclR, PadR), quorum sensing (TetR, LysR, IclR); sporulation (IclR); detoxification (PadR); salt stress response (DegU); and copper stress response (CutR) (Huillet et al., 2006; Molina-Henares et al., 2006; Maddocks and Oyston, 2008; Fibriansah et al., 2012; Rademacher and Masepohl, 2012; Cuthbertson and Nodwell, 2013; Rodríguez et al., 2013; Tian et al., 2014; Hoffmann and Bremer, 2016).

For example, a gene encoding a LysR family transcriptional regulator, that is present uniquely in the SM18 genome, is located upstream of a gene which appears to encode a Beta-lactamase enzyme family protein, which are enzymes that provide mechanisms of resistance to  $\beta$ -lactam antibiotics (Majiduddin et al., 2002; Naas et al., 2017). In addition, a gene encoding a IclR family transcriptional regulator that is present in both the SM17 and SM18 genomes, is located upstream of a *proP* gene, which potentially encodes a proline/betaine transporter and, which previously mentioned, could be related to osmotic regulation in these organisms (MacMillan et al., 1999; Roessler and Muller, 2001; Burg and Ferraris, 2008). While a gene encoding a transcriptional regulator PadR-like family protein which is present in both the SM17 and SM18 genomes; is located upstream of a gene coding an ABC transporter, which as previously mentioned, could be involved with nutrient acquisition, resistance to toxic molecules, or in maintaining osmotic balance in these isolates (Wood, 2007; Fan et al., 2013; Greene et al., 2018; Teichmann et al., 2018).

#### 4.5.3.5. Genomic evolution through horizontal gene transfer

Horizontal gene transfer (HGT) is an important mechanism in bacterial genome evolution, and commonly involves the acquisition of mobile genetic elements (MGEs) (Bellanger et al., 2014). Previous studies have reported that the

genomes of symbiotic bacteria – including sponge symbionts – possess a higher number of MGEs than those of free-living microorganisms (Thomas et al., 2010; Fan et al., 2012, 2013). It has been proposed that MGEs play a crucial role in co-evolution with the host and convergent evolution of marine sponge symbiotic communities in a number of ways, such as enabling the members of the symbiotic community to share important traits for niche adaptation (Fan et al., 2012), such as for example genes related to stress tolerance, antibiotics resistance, and nutrient acquisition. In addition, the MGEs can function in the deactivation or removal of non-essential genes, such as those that are only required by free-living bacteria, or those related to functions that are already being performed by other members of the symbiotic community (Fan et al., 2012). Two genes encoding transposases were found in the ENA gene pool (Table 3), indicating that they may be involved in HGT events and co-evolution between the marine sponge isolates SM17 and SM18. Also, the three plasmids that were identified in SM17 (Table 1B and Figure 1), which are absent in its terrestrial counterpart *S. albus* J1074, provide additional evidence of potential genomic evolution through transferable elements occurring within the marine sponge microbiota.

The high filter feeding rates of sponges mean that they are likely to be exposed to phage attack from the plankton, and that bacterial sponge symbionts may be subjected to phage-mediated transduction which can lead to cell lysis (Thomas et al., 2010). Therefore, it might be expected that sponge bacterial symbiotic communities would require defense mechanisms to protect themselves from foreign DNA, such as restriction modification (R-M) systems and toxin-antitoxin (T-A) systems (Fan et al., 2012; Horn et al., 2016; Slaby et al., 2017). R-M systems are also linked to MGEs in that they can be transferred via the MGEs, or they can act as MGEs in transposon-like structures (Furuta and Kobayashi, 2013). In the ENA gene pool, we identified one restriction endonuclease that could be part

of a transferrable R-M system and one T-A system gene from the RelE family in SM17 and SM18 (Table 3). This further highlights the possibility that HGT events may be occurring between the sponge-derived isolates and the possibility of shared niche adaptations between them, and also the requirement for defense mechanisms against foreign DNA in the symbiotic bacteria. Importantly, T-A systems have also been proposed to provide mechanisms to cope with stress – such as nutrient stress – by either programmed cell death or by inducing bacteriostasis, which may be another important role played by the T-A systems in symbiotic communities in oligotrophic environments (Van Melderren, 2010; de Goeij et al., 2013).

#### 4.5.3.6. Eukaryotic-like proteins and potential host interaction

Metagenomic and genomic studies have reported that bacterial symbionts contain a large number of genes encoding for eukaryotic-like proteins (ELPs) (Reynolds and Thomas, 2016). ELPs contain repeat domains that are commonly found in eukaryotic proteins, such as tetratricopeptide repeats (TPRs), and are believed to play an important role in symbiotic relationships, by mediating protein-protein interactions for a range of cellular proteins (Thomas et al., 2010; Li et al., 2015; Reynolds and Thomas, 2016). These ELPs may have a broader function in mediating bacterial-sponge interactions and may modulate the host's behavior (Li et al., 2015; Reynolds and Thomas, 2016). The ENA gene pool contained a tetratricopeptide repeat-containing protein, which is a class of ELP that has been proposed to function as a means for symbiotic bacteria to avoid digestion, or as a mechanism for the sponge to distinguish between food and symbionts (Thomas et al., 2010). The fact that the relatively phylogenetically distant SM17 and SM18 isolates possess orthologs of the same TPR, while their closest terrestrial relatives do not; suggests that this protein may indeed play a role in the symbiotic interactions between these bacteria and their sponge host *Haliclona simulans*.

#### 4.5.4. ENA gene pool genes commonly present in other environmental

##### *Streptomyces* isolates

In a similar fashion to the aforementioned analysis of orthologous genes, an additional analysis was performed, including the genomes from the other isolates previously determined to belong to the SM17 and SM18 phylogenetic clades (Figure 2). The aim was to assess whether genes present in the SM17 and SM18's ENA gene pool are also present in other closely related relatives derived from other diverse environments, given the possibility that they may possess adaptations to their particular environmental niches that overlap with those identified in our marine sponge-associated SM17 and SM18 strains.

In the previously identified SM17 clade (Figure 2), in addition to its closely related terrestrial type-strain J1074, the clade also included the environmental isolates *Streptomyces albidoflavus* SM254, which was isolated from copper-rich subsurface fluids within an iron mine (Badalamenti et al., 2016); *Streptomyces sampsonii* KJ40, which was isolated from rhizosphere soil in a poplar plantation (Li et al., 2018); *Streptomyces koyangensis* VK-A60T, which was isolated from rhizosphere soil in a radish plantation (Lee et al., 2005); and *Streptomyces* sp. CLI2509, which is a fungus-derived isolate (Wyche et al., 2017). It is important to note that the SM17 clade also included the *Streptomyces* sp. FR-008 strain, however, this strain was not included in the analysis since it does not appear to be an environmental isolate, and it is a product of protoplast breeding of strains with little information in the literature regarding their isolation source (NCBI BioSample: SAMN03120580). The SM18 clade (Figure 2), in addition to its closely related terrestrial type-strain ATCC 33331, also included the environmental isolates *Streptomyces* sp. PAMC26508, which is an endosymbiotic bacterium isolated from the Antarctic lichen *Cladonia borealis* (Shin et al., 2013); *Streptomyces* sp. S501, isolated in sediment from a seaside wetland (NCBI BioSample: SAMN10144670);

2093 and *Streptomyces* sp. SirexAA-E, isolated from an insect/microbe symbiotic  
2094 community (Bianchetti et al., 2013).

2095 Interestingly, the majority of the genes present in the ENA gene pool were  
2096 also present in the genomes of the other isolates. This is perhaps not surprising  
2097 given the potential similarity in environmental stresses that these isolates may  
2098 encounter, as the marine sponge-associated SM17 and SM18 strains; since they  
2099 were all isolated from either (1) symbiotic communities, (2) high osmotic pressure  
2100 environments and/or 3) aquatic environments. For example, the aforementioned  
2101 *nuo* operon genes (Figure 7); potentially involved in adaptation to osmotic stress,  
2102 are also present in the KJ40, the VK-A60T, and SM254 strains from the SM17 clade  
2103 (Figure 2). It is well documented that osmoadaptation is an important trait  
2104 possessed by rhizosphere-derived bacteria, since water uptake and exclusion of  
2105 solutes such as Na<sup>+</sup> and Cl<sup>-</sup> by plants roots are likely to induce changes in  
2106 osmolarity (Miller and Wood, 1996; Qurashi and Sabri, 2011), and for that reason  
2107 salt-tolerant bacterium are commonly isolated from plant rhizospheres (Yuwono,  
2108 2005; Qurashi and Sabri, 2011). Thus, it is reasonable to assume that the presence of  
2109 the *nuo* operon genes in the KJ40 and in the VK-A60T strains, both rhizosphere-  
2110 derived isolates, may also be related to an increased resistance to osmotic stress, as  
2111 it also seems to be the case to our marine sponge-derived isolates. Likewise, it is  
2112 also possible that the SM254 strain, isolated from copper-rich subsurface fluids in  
2113 an iron mine, will be exposed to osmotic stress and hence require appropriate  
2114 adaptations to these conditions. Hence, it is plausible that the genes encoded in the  
2115 *nuo* operon are not an adaptive response that is exclusively employed by some  
2116 marine bacteria, as previously suggested (Penn and Jensen, 2012; Ian et al., 2014;  
2117 Sun et al., 2018), but rather a more general mechanism of osmoadaptation that may  
2118 be employed by bacteria in other environments as well.

2119            Similarly, *proP* gene homologs were also present in all of the other genomes  
2120 analyzed, with exception to the fungus-derived CLI2905 strain from the SM17  
2121 clade (Figure 2). Thus, given as has been previously discussed, that ProP acts both  
2122 as an osmoregulator and as an osmosensor, together with transporting compatible  
2123 solutes in *E. coli*; it may also be related to osmoadaptation in these isolates. These  
2124 observations further highlight the potential adaptations which have been proposed  
2125 in the ENA gene pool, that may be present in these other closely related relatives  
2126 derived from other diverse environments, which may overlap with those identified  
2127 in our marine sponge-associated SM17 and SM18 strains.

## 5. Conclusion

The *Streptomyces* genus is exceptionally important when it comes to the identification and production of bioactive molecules, but those derived from the marine environment are currently particularly not well characterized. This study provides novel insights into possible ENAs employed by *Streptomyces* spp. isolated from marine sponges, and how these are potentially linked to diverse secondary metabolite biosynthesis. By providing high quality genomic information for the SM17 and SM18 strains isolated from *Haliclona simulans*, which have been previously shown to have antimicrobial activity against important pathogens, we were able to perform several comparative analyses with their terrestrial counterparts *S. albus* J1074 and *S. pratensis* ATCC 33331. The genomic analyses identified a diversity of putative smBGCs, which could potentially explain the previously determined antimicrobial activities reported for these marine isolates, such as smBGCs potentially encoding the production of candicidin, antimycin, SGR PTMs, surugamides, and mannopeptimycin, in SM17; and smBGCs potentially encoding the production of bafilomycin, SGR PTMs, curamycin, and caboxamycin, in SM18. Several smBGCs appear to be unique in the marine isolates in comparison to their terrestrial counterparts, which is particularly true in the case of the *Streptomyces* sp. SM18 isolate, when compared to *S. pratensis* ATCC 33331. Interestingly, while SM18 contains smBGCs encoding the production of siderophores such as coelichelin and mirubactin, it lacks the smBGC encoding the production of desferrioxamines; which is to our knowledge the first report of a *Streptomyces* isolate lacking this capacity. Comparative genomics analysis allowed us to identify genes that could be involved in mechanisms that may be relevant for their adaptation to their particular environmental niche, including resistance to osmotic stress; transcriptional regulation; symbiotic interactions; antimicrobial compounds production and resistance; ABC transporters; and HGT and other



2155 potential defense-related features. Expanding on the genetic knowledge of these  
2156 organisms and their underlying mechanisms of adaptability is important, in not  
2157 only allowing us to gain a better understanding of marine bacteria and their  
2158 evolution, but also in helping with the discovery of potential new bioactive small  
2159 molecules and in how to potentially manipulate and optimize their production.

2160    **6.    References**

- 2161    Abdelmohsen, U. R., Bayer, K., and Hentschel, U. (2014). Diversity, abundance and  
 2162    natural products of marine sponge-associated actinomycetes. *Nat. Prod. Rep.*  
 2163    31, 381–399. doi:10.1039/C3NP70111E.
- 2164    Adnani, N., Rajski, S. R., and Bugni, T. S. (2017). Symbiosis-inspired approaches to  
 2165    antibiotic discovery. *Nat. Prod. Rep.* 34, 784–814. doi:10.1039/C7NP00009J.
- 2166    Almeida, E. L., Margassery, L. M., Kennedy, J., and Dobson, A. D. W. (2018). Draft  
 2167    Genome Sequence of the Antimycin-Producing Bacterium *Streptomyces* sp.  
 2168    Strain SM8, Isolated from the Marine Sponge *Haliclona simulans*. *Genome*  
 2169    *Announc.* 6, e01535-17. doi:10.1128/genomeA.01535-17.
- 2170    Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., and Suzuki, M. T.  
 2171    (2017). Multiple *Streptomyces* species with distinct secondary metabolomes  
 2172    have identical 16S rRNA gene sequences. *Sci. Rep.* 7, 11089. doi:10.1038/s41598-  
 2173    017-11363-1.
- 2174    Ay, H., Nouiou, I., del Carmen Montero-Calasanz, M., Klenk, H.-P., Isik, K., Cetin,  
 2175    D., et al. (2018). *Streptomyces sediminis* sp. nov. isolated from crater lake  
 2176    sediment. *Antonie Van Leeuwenhoek* 111, 493–500. doi:10.1007/s10482-017-  
 2177    0970-z.
- 2178    Badalamenti, J. P., Erickson, J. D., and Salomon, C. E. (2016). Complete Genome  
 2179    Sequence of *Streptomyces albus* SM254, a Potent Antagonist of Bat White-Nose  
 2180    Syndrome Pathogen *Pseudogymnoascus destructans*. *Genome Announc.* 4.  
 2181    doi:10.1128/genomeA.00290-16.
- 2182    Baker, P. W., Kennedy, J., Dobson, A. D. W., and Marchesi, J. R. (2009). Phylogenetic  
 2183    Diversity and Antimicrobial Activities of Fungi Associated with *Haliclona*  
 2184    *simulans* Isolated from Irish Coastal Waters. *Mar. Biotechnol.* 11, 540–547.  
 2185    doi:10.1007/s10126-008-9169-7.
- 2186    Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P., and Marth, G. T.  
 2187    (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files.  
 2188    *Bioinformatics* 27, 1691–1692. doi:10.1093/bioinformatics/btr174.
- 2189    Barra, L., Barac, P., König, G. M., Crüsemann, M., and Dickschat, J. S. (2017).  
 2190    Volatiles from the fungal microbiome of the marine sponge *Callyspongia* cf.  
 2191    *flammea*. *Org. Biomol. Chem.* 15, 7411–7421. doi:10.1039/C7OB01837A.
- 2192    Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative  
 2193    and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS*  
 2194    *Microbiol. Rev.* 38, 720–760. doi:10.1111/1574-6976.12058.

- 2195 Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D.,  
2196 et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47.  
2197 doi:10.1093/nar/gkx1094.
- 2198 Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R.,  
2199 James, K. D., et al. (2002). Complete genome sequence of the model  
2200 actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147.  
2201 doi:10.1038/417141a.
- 2202 Bianchetti, C. M., Harmann, C. H., Takasuka, T. E., Hura, G. L., Dyer, K., and Fox, B.  
2203 G. (2013). Fusion of Dioxygenase and Lignin-binding Domains in a Novel  
2204 Secreted Enzyme from Cellulolytic *Streptomyces* sp. SirexAA-E. *J. Biol. Chem.*  
2205 288, 18574–18587. doi:10.1074/jbc.M113.475848.
- 2206 Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019).  
2207 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.  
2208 *Nucleic Acids Res.* doi:10.1093/nar/gkz310.
- 2209 Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al.  
2210 (2017). antiSMASH 4.0—improvements in chemistry prediction and gene  
2211 cluster boundary identification. *Nucleic Acids Res.* 45, W36–W41.  
2212 doi:10.1093/nar/gkx319.
- 2213 Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., and Schulz-Vogt, H. N.  
2214 (2013). The genus *Pseudovibrio* contains metabolically versatile bacteria  
2215 adapted for symbiosis. *Environ. Microbiol.* 15, 2095–2113. doi:10.1111/1462-  
2216 2920.12123.
- 2217 Bowman, E. J., Siebers, A., and Altendorf, K. (1988). Bafilomycins: a class of  
2218 inhibitors of membrane ATPases from microorganisms, animal cells, and plant  
2219 cells. *Proc. Natl. Acad. Sci.* 85, 7972–7976. doi:10.1073/pnas.85.21.7972.
- 2220 Burg, M. B., and Ferraris, J. D. (2008). Intracellular organic osmolytes: function and  
2221 regulation. *J. Biol. Chem.* 283, 7309–13. doi:10.1074/jbc.R700042200.
- 2222 Campelo, A. B., and Gil, J. A. (2002). The candicidin gene cluster from *Streptomyces*  
2223 *griseus* IMRU 3570. *Microbiology* 148, 51–59. doi:10.1099/00221287-148-1-51.
- 2224 Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009).  
2225 DNAPlotter: circular and linear interactive genome visualization.  
2226 *Bioinformatics* 25, 119–120. doi:10.1093/bioinformatics/btn578.
- 2227 Ceapă, C. D., Vázquez-Hernández, M., Rodríguez-Luna, S. D., Cruz Vázquez, A. P.,  
2228 Jiménez Suárez, V., Rodríguez-Sanoja, R., et al. (2018). Genome mining of  
2229 *Streptomyces scabrissporus* NF3 reveals symbiotic features including genes  
2230 related to plant interactions. *PLoS One* 13, e0192618.

- doi:10.1371/journal.pone.0192618.
- Chater, K. F., and Wilde, L. C. (1976). Restriction of a bacteriophage of *Streptomyces albus* G involving endonuclease *Sa*II. *J. Bacteriol.* 128, 644–50.
- Chater, K. F., and Wilde, L. C. (1980). *Streptomyces albus* G Mutants Defective in the *Sa*IGI Restriction-Modification System. *Microbiology* 116, 323–334. doi:10.1099/00221287-116-2-323.
- Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., et al. (2003). Organizational and Mutational Analysis of a Complete FR-008/Candididin Gene Cluster Encoding a Structurally Related Polyene Complex. *Chem. Biol.* 10, 1065–1076. doi:10.1016/J.CHEMBIOL.2003.10.007.
- Cheng, C., Othman, E., Stopper, H., Edrada-Ebel, R., Hentschel, U., Abdelmohsen, U., et al. (2017). Isolation of Petrocidin A, a New Cytotoxic Cyclic Dipeptide from the Marine Sponge-Derived Bacterium *Streptomyces* sp. SBT348. *Mar. Drugs* 15, 383. doi:10.3390/md15120383.
- Cheng, Y.-Q., Tang, G.-L., and Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci.* 100, 3149–3154. doi:10.1073/pnas.0537286100.
- Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., et al. (2006a). Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome. *Mol. Biol. Evol.* 23, 2361–2369. doi:10.1093/molbev/msl108.
- Choulet, F., Gallois, A., Aigle, B., Mangenot, S., Gerbaud, C., Truong, C., et al. (2006b). Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. *J. Bacteriol.* 188, 6599–610. doi:10.1128/JB.00734-06.
- Čihák, M., Kameník, Z., Šmídová, K., Bergman, N., Benada, O., Kofroňová, O., et al. (2017). Secondary Metabolites Produced during the Germination of *Streptomyces coelicolor*. *Front. Microbiol.* 8, 2495. doi:10.3389/fmicb.2017.02495.
- Clay, K. (2014). Defensive symbiosis: a microbial perspective. *Funct. Ecol.* 28, 293–298. doi:10.1111/1365-2435.12258.
- Collins, B., Guinane, C. M., Cotter, P. D., Hill, C., and Ross, R. P. (2012). Assessing the contributions of the LiaS histidine kinase to the innate resistance of *Listeria monocytogenes* to nisin, cephalosporins, and disinfectants. *Appl. Environ. Microbiol.* 78, 2923–9. doi:10.1128/AEM.07402-11.
- Cruz-Morales, P., Ramos-Aboites, H. E., Licona-Cassani, C., Selem-Mójica, N., Mejía-Ponce, P. M., Souza-Saldívar, V., et al. (2017). Actinobacteria

- 2267 phylogenomics, selective isolation from an iron oligotrophic environment and  
 2268 siderophore functional characterization, unveil new desferrioxamine traits.  
 2269 *FEMS Microbiol. Ecol.* 93. doi:10.1093/femsec/fix086.
- 2270 Cuthbertson, L., and Nodwell, J. R. (2013). The TetR Family of Regulators.  
 2271 *Microbiol. Mol. Biol. Rev.* 77, 440–475. doi:10.1128/MMBR.00018-13.
- 2272 Dalisay, D. S., Williams, D. E., Wang, X. L., Centko, R., Chen, J., and Andersen, R. J.  
 2273 (2013). Marine Sediment-Derived *Streptomyces* Bacteria from British  
 2274 Columbia, Canada Are a Promising Microbiota Resource for the Discovery of  
 2275 Antimicrobial Natural Products. *PLoS One* 8, e77078.  
 2276 doi:10.1371/journal.pone.0077078.
- 2277 Daniel-Ivad, M., Pimentel-Elardo, S., and Nodwell, J. R. (2018). Control of  
 2278 Specialized Metabolism by Signaling and Transcriptional Regulation:  
 2279 Opportunities for New Platforms for Drug Discovery? *Annu. Rev. Microbiol.*  
 2280 72, 25–48. doi:10.1146/annurev-micro-022618-042458.
- 2281 de Goeij, J. M., van Oevelen, D., Vermeij, M. J. A., Osinga, R., Middelburg, J. J., de  
 2282 Goeij, A. F. P. M., et al. (2013). Surviving in a marine desert: the sponge loop  
 2283 retains resources within coral reefs. *Science* 342, 108–10.  
 2284 doi:10.1126/science.1241981.
- 2285 Dusa, A. (2018). venn: Draw Venn Diagrams. Available at: [https://cran.r-](https://cran.r-project.org/package=venn)  
 2286 [project.org/package=venn](https://cran.r-project.org/package=venn).
- 2287 Engl, T., Kroiss, J., Kai, M., Nechitaylo, T. Y., Svatoš, A., and Kaltenpoth, M. (2018).  
 2288 Evolutionary stability of antibiotic protection in a defensive symbiosis. *Proc.*  
 2289 *Natl. Acad. Sci.* 115, E2020–E2029. doi:10.1073/pnas.1719797115.
- 2290 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al.  
 2291 (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47,  
 2292 D427–D432. doi:10.1093/nar/gky995.
- 2293 Fan, L., Liu, M., Simister, R., Webster, N. S., and Thomas, T. (2013). Marine  
 2294 microbial symbiosis heats up: the phylogenetic and functional response of a  
 2295 sponge holobiont to thermal stress. *ISME J.* 7, 991–1002.  
 2296 doi:10.1038/ismej.2012.165.
- 2297 Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S., et al. (2012).  
 2298 Functional equivalence and evolutionary convergence in complex  
 2299 communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci.* 109, E1878–  
 2300 E1887. doi:10.1073/pnas.1203287109.
- 2301 Feklistov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial Sigma  
 2302 Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev.*

- 2303 *Microbiol.* 68, 357–376. doi:10.1146/annurev-micro-092412-155737.
- 2304 Fibriansah, G., Kovács, Á. T., Pool, T. J., Boonstra, M., Kuipers, O. P., and  
 2305 Thunnissen, A.-M. W. H. (2012). Crystal Structures of Two Transcriptional  
 2306 Regulators from *Bacillus cereus* Define the Conserved Structural Features of a  
 2307 PadR Subfamily. *PLoS One* 7, e48015. doi:10.1371/journal.pone.0048015.
- 2308 Flemer, B., Kennedy, J., Margassery, L. M., Morrissey, J. P., O’Gara, F., and Dobson,  
 2309 A. D. W. (2012). Diversity and antimicrobial activities of microbes from two  
 2310 Irish marine sponges, *Suberites carnosus* and *Leucosolenia* sp. *J. Appl.*  
 2311 *Microbiol.* 112, 289–301. doi:10.1111/j.1365-2672.2011.05211.x.
- 2312 Fuerst, J. A. (2014). Diversity and biotechnological potential of microorganisms  
 2313 associated with marine sponges. *Appl. Microbiol. Biotechnol.* 98, 7331–7347.  
 2314 doi:10.1007/s00253-014-5861-x.
- 2315 Furuta, Y., and Kobayashi, I (2013). “Restriction-modification systems as mobile  
 2316 epigenetic elements,” in *Madame Curie Bioscience Database* (Austin, TX, USA:  
 2317 Landes Bioscience). Available at:  
 2318 <https://www.ncbi.nlm.nih.gov/books/NBK63963/>
- 2319 Gadagkar, S. R., Rosenberg, M. S., and Kumar, S. (2005). Inferring species  
 2320 phylogenies from multiple genes: Concatenated sequence tree versus  
 2321 consensus gene tree. *J. Exp. Zool. Part B Mol. Dev. Evol.* 304B, 64–74.  
 2322 doi:10.1002/jez.b.21026.
- 2323 Galmarini, O. L., and Deulofeu, V. (1961). Curamycin—I: Isolation and  
 2324 characterization of some hydrolysis products. *Tetrahedron* 15, 76–86.  
 2325 doi:10.1016/0040-4020(61)80010-0.
- 2326 Gomez-Escribano, J. P., Castro, J. F., Razmilic, V., Chandra, G., Andrews, B., Asenjo,  
 2327 J. A., et al. (2015). The *Streptomyces leeuwenhoekii* genome: *de novo*  
 2328 sequencing and assembly in single contigs of the chromosome, circular  
 2329 plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics* 16, 485.  
 2330 doi:10.1186/s12864-015-1652-8.
- 2331 González-Pastor, J. E., Hobbs, E. C., and Losick, R. (2003). Cannibalism by  
 2332 sporulating bacteria. *Science* 301, 510–3. doi:10.1126/science.1086462.
- 2333 Goodfellow, M., and Fiedler, H. P. (2010). A guide to successful bioprospecting:  
 2334 Informed by actinobacterial systematics. *Antonie van Leeuwenhoek, Int. J. Gen.*  
 2335 *Mol. Microbiol.* 98, 119–142. doi:10.1007/s10482-010-9460-2.
- 2336 Goshi, K., Uchida, T., Lezhava, A., Yamasaki, M., Hiratsu, K., Shinkawa, H., et al.  
 2337 (2002). Cloning and Analysis of the Telomere and Terminal Inverted Repeat of  
 2338 the Linear Chromosome of *Streptomyces griseus*. *J. Bacteriol.* 184, 3411–3415.

- 2339 doi:10.1128/JB.184.12.3411-3415.2002.
- 2340 Greene, N. P., Kaplan, E., Crow, A., and Koronakis, V. (2018). Antibiotic Resistance  
2341 Mediated by the MacB ABC Transporter Family: A Structural and Functional  
2342 Perspective. *Front. Microbiol.* 9, 950. doi:10.3389/fmicb.2018.00950.
- 2343 Gromek, S. M., Suria, A. M., Fullmer, M. S., Garcia, J. L., Gogarten, J. P., Nyholm, S.  
2344 V., et al. (2016). *Leisingera* sp. JC1, a Bacterial Isolate from Hawaiian Bobtail  
2345 Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios. *Front.*  
2346 *Microbiol.* 7, 1342. doi:10.3389/fmicb.2016.01342.
- 2347 Guo, P., Cheng, Q., Xie, P., Fan, Y., Jiang, W., and Qin, Z. (2011). Characterization of  
2348 the multiple CRISPR loci on *Streptomyces* linear plasmid pSHK1. *Acta*  
2349 *Biochim. Biophys. Sin. (Shanghai)*. 43, 630–639. doi:10.1093/abbs/gmr052.
- 2350 Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality  
2351 assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075.  
2352 doi:10.1093/bioinformatics/btt086.
- 2353 Hassan, S. S. ul, Anjum, K., Abbas, S. Q., Akhter, N., Shagufta, B. I., Shah, S. A. A.,  
2354 et al. (2017). Emerging biopharmaceuticals from marine actinobacteria.  
2355 *Environ. Toxicol. Pharmacol.* 49, 34–47. doi:10.1016/j.etap.2016.11.015.
- 2356 Hider, R. C., and Kong, X. (2010). Chemistry and biology of siderophores. *Nat.*  
2357 *Prod. Rep.* 27, 637. doi:10.1039/b906679a.
- 2358 Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., and Leblond, P. (2018).  
2359 Genome plasticity is governed by double strand break DNA repair in  
2360 *Streptomyces*. *Sci. Rep.* 8, 5272. doi:10.1038/s41598-018-23622-w.
- 2361 Hoffmann, T., and Bremer, E. (2016). “Management of Osmotic Stress by *Bacillus*  
2362 *Subtilis*: Genetics and Physiology,” in *Stress and Environmental Regulation of*  
2363 *Gene Expression and Adaptation in Bacteria* (Hoboken, NJ, USA: John Wiley &  
2364 Sons, Inc.), 657–676. doi:10.1002/9781119004813.ch63.
- 2365 Hohmann, C., Schneider, K., Bruntner, C., Irran, E., Nicholson, G., Bull, A. T., et al.  
2366 (2009). Caboxamycin, a new antibiotic of the benzoxazole family produced by  
2367 the deep-sea strain *Streptomyces* sp. NTK 937. *J. Antibiot. (Tokyo)*. 62, 99–104.  
2368 doi:10.1038/ja.2008.24.
- 2369 Hoppers, A., Stoudenmire, J., Wu, S., and Lopanik, N. B. (2015). Antibiotic activity  
2370 and microbial community of the temperate sponge, *Haliclona* sp. *J. Appl.*  
2371 *Microbiol.* 118, 419–430. doi:10.1111/jam.12709.
- 2372 Horn, H., Slaby, B. M., Jahn, M. T., Bayer, K., Moitinho-Silva, L., Förster, F., et al.  
2373 (2016). An Enrichment of CRISPR and Other Defense-Related Features in

- 2374 Marine Sponge-Associated Microbial Metagenomes. *Front. Microbiol.* 7, 1751.  
2375 doi:10.3389/fmicb.2016.01751.
- 2376 Hoz, J. F.-D. la, Méndez, C., Salas, J. A., and Olano, C. (2017). Novel Bioactive  
2377 Paulomycin Derivatives Produced by *Streptomyces albus* J1074. *Molecules* 22,  
2378 1758. doi:10.3390/molecules22101758.
- 2379 Huillet, E., Velge, P., Vallaëys, T., and Pardon, P. (2006). LadR, a new PadR-related  
2380 transcriptional regulator from *Listeria monocytogenes*, negatively regulates  
2381 the expression of the multidrug efflux pump MdrL. *FEMS Microbiol. Lett.* 254,  
2382 87–94. doi:10.1111/j.1574-6968.2005.00014.x.
- 2383 Hwang, K.-S., Kim, H. U., Charusanti, P., Palsson, B. Ø., and Lee, S. Y. (2014).  
2384 Systems biology and biotechnology of *Streptomyces* species for the production  
2385 of secondary metabolites. *Biotechnol. Adv.* 32, 255–268.  
2386 doi:10.1016/j.biotechadv.2013.10.008.
- 2387 Ian, E., Malko, D. B., Sekurova, O. N., Bredholt, H., Rückert, C., Borisova, M. E., et  
2388 al. (2014). Genomics of Sponge-Associated *Streptomyces* spp. Closely Related  
2389 to *Streptomyces albus* J1074: Insights into Marine Adaptation and Secondary  
2390 Metabolite Biosynthesis Potential. *PLoS One* 9, e96719.  
2391 doi:10.1371/journal.pone.0096719.
- 2392 Indraningrat, A., Smidt, H., and Sipkema, D. (2016). Bioprospecting Sponge-  
2393 Associated Microbes for Antimicrobial Compounds. *Mar. Drugs* 14, 87.  
2394 doi:10.3390/md14050087.
- 2395 Iniyan, A. M., Mary, T. R. J., Joseph, F.-J. R. S., Kannan, R. R., and Vincent, S. G. P.  
2396 (2016). Cell wall distracting anti-Methicillin-resistant *Staphylococcus aureus*  
2397 compound PVI331 from a marine sponge associated *Streptomyces*. *J. Appl.*  
2398 *Biomed.* 14, 273–283. doi:10.1016/j.jab.2016.04.003.
- 2399 Jackson, S., Crossman, L., Almeida, E., Margassery, L., Kennedy, J., and Dobson, A.  
2400 (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene  
2401 Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates.  
2402 *Mar. Drugs* 16, 67. doi:10.3390/md16020067.
- 2403 Jin, J., Yang, X., Liu, T., Xiao, H., Wang, G., Zhou, M., et al. (2018). Fluostatins M–Q  
2404 Featuring a 6-5-6-6 Ring Skeleton and High Oxidized A-Rings from Marine  
2405 *Streptomyces* sp. PKU-MA00045. *Mar. Drugs* 16, 87. doi:10.3390/md16030087.
- 2406 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden,  
2407 T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–  
2408 W9. doi:10.1093/nar/gkn201.
- 2409 Kallifidas, D., Jiang, G., Ding, Y., and Luesch, H. (2018). Rational engineering of



- 2410 *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene  
2411 clusters. *Microb. Cell Fact.* 17, 25. doi:10.1186/s12934-018-0874-2.
- 2412 Kamjam, M., Sivalingam, P., Deng, Z., and Hong, K. (2017). Deep Sea  
2413 Actinomycetes and Their Secondary Metabolites. *Front. Microbiol.* 8, 760.  
2414 doi:10.3389/fmicb.2017.00760.
- 2415 Karimi, E., Slaby, B. M., Soares, A. R., Blom, J., Hentschel, U., and Costa, R. (2018).  
2416 Metagenomic binning reveals versatile nutrient cycling and distinct adaptive  
2417 features in alphaproteobacterial symbionts of marine sponges. *FEMS*  
2418 *Microbiol. Ecol.* 94. doi:10.1093/femsec/fiy074.
- 2419 Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment  
2420 Software Version 7: Improvements in Performance and Usability. *Mol. Biol.*  
2421 *Evol.* 30, 772–780. doi:10.1093/molbev/mst010.
- 2422 Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., et al. (2009).  
2423 Isolation and Analysis of Bacteria with Antimicrobial Activities from the  
2424 Marine Sponge *Haliclona simulans* Collected from Irish Waters. *Mar.*  
2425 *Biotechnol.* 11, 384–396. doi:10.1007/s10126-008-9154-1.
- 2426 Kiran, G. S., Sekar, S., Ramasamy, P., Thinesh, T., Hassan, S., Lipton, A. N., et al.  
2427 (2018). Marine sponge microbial association: Towards disclosing unique  
2428 symbiotic interactions. *Mar. Environ. Res.* 140, 169–179.  
2429 doi:10.1016/j.marenvres.2018.04.017.
- 2430 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M.  
2431 (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer  
2432 weighting and repeat separation. *Genome Res.* 27, 722–736.  
2433 doi:10.1101/gr.215087.116.
- 2434 Krämer, R. (2010). Bacterial stimulus perception and signal transduction: Response  
2435 to osmotic stress. *Chem. Rec.* 10, 217–229. doi:10.1002/tcr.201000005.
- 2436 Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X:  
2437 Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol.*  
2438 *Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.
- 2439 Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al.  
2440 (2004). Versatile and open software for comparing large genomes. *Genome*  
2441 *Biol.* 5, R12. doi:10.1186/gb-2004-5-2-r12.
- 2442 Kwan, J. C., Tianero, M. D. B., Donia, M. S., Wyche, T. P., Bugni, T. S., and Schmidt,  
2443 E. W. (2014). Host Control of Symbiont Natural Product Chemistry in Cryptic  
2444 Populations of the Tunicate *Lissoclinum patella*. *PLoS One* 9, e95850.  
2445 doi:10.1371/journal.pone.0095850.

- 2446 Lee, J. Y., Lee, J. Y., Jung, H. W., and Hwang, B. K. (2005). *Streptomyces*  
2447 *koyangensis* sp. nov., a novel actinomycete that produces 4-phenyl-3-butenic  
2448 acid. *Int. J. Syst. Evol. Microbiol.* 55, 257–262. doi:10.1099/ijs.0.63168-0.
- 2449 Lee, L.-H., Chan, K.-G., Stach, J., Wellington, E. M. H., and Goh, B.-H. (2018).  
2450 Editorial: The Search for Biological Active Agent(s) From Actinobacteria. *Front.*  
2451 *Microbiol.* 9, 824. doi:10.3389/fmicb.2018.00824.
- 2452 Lee, S. -o., Kato, J., Takiguchi, N., Kuroda, A., Ikeda, T., Mitsutani, A., et al. (2000).  
2453 Involvement of an Extracellular Protease in Algicidal Activity of the Marine  
2454 Bacterium *Pseudoalteromonas* sp. Strain A28. *Appl. Environ. Microbiol.* 66,  
2455 4334–4339. doi:10.1128/AEM.66.10.4334-4339.2000.
- 2456 Li, H.-J., Tang, B.-L., Shao, X., Liu, B.-X., Zheng, X.-Y., Han, X.-X., et al. (2016).  
2457 Characterization of a New S8 serine Protease from Marine Sedimentary  
2458 *Photobacterium* sp. A5–7 and the Function of Its Protease-Associated Domain.  
2459 *Front. Microbiol.* 7, 2016. doi:10.3389/fmicb.2016.02016.
- 2460 Li, Q., Yi, L., Marek, P., and Iverson, B. L. (2013). Commercial proteases: Present  
2461 and future. *FEBS Lett.* 587, 1155–1163. doi:10.1016/j.febslet.2012.12.019.
- 2462 Li, S., Zhang, B., Zhu, H., and Zhu, T. (2018). Cloning and Expression of the  
2463 Chitinase Encoded by ChiKJ406136 from *Streptomyces Sampsonii* (Millard  
2464 & Burr) Waksman KJ40 and Its Antifungal Effect. *Forests* 9, 699.  
2465 doi:10.3390/f9110699.
- 2466 Li, Z.-Y., Wang, Y.-Z., He, L.-M., and Zheng, H.-J. (2015). Metabolic profiles of  
2467 prokaryotic and eukaryotic communities in deep-sea sponge *Neamphius*  
2468 *huxleyi* indicated by metagenomics. *Sci. Rep.* 4, 3895. doi:10.1038/srep03895.
- 2469 Lim, B., and Lee, K. (2015). Stability of the osmoregulated promoter-derived *proP*  
2470 mRNA is posttranscriptionally regulated by RNase III in *Escherichia coli*. *J.*  
2471 *Bacteriol.* 197, 1297–305. doi:10.1128/JB.02460-14.
- 2472 Lin, Y., Hahn, M.-Y., Roe, J.-H., Huang, T.-W., Tsai, H.-H., Lin, Y.-F., et al. (2009).  
2473 *Streptomyces* Telomeres Contain a Promoter. *J. Bacteriol.* 191, 773–781.  
2474 doi:10.1128/JB.01299-08.
- 2475 Linares, J. F., Gustafsson, I., Baquero, F., and Martinez, J. L. (2006). Antibiotics as  
2476 intermicrobial signaling agents instead of weapons. *Proc. Natl. Acad. Sci.* 103,  
2477 19484–19489. doi:10.1073/pnas.0608949103.
- 2478 Liu, C., Jiang, Y., Lei, H., Chen, X., Ma, Q., Han, L., et al. (2017). Four New  
2479 Nanaomycins Produced by *Streptomyces hebeiensis* Derived from Lichen.  
2480 *Chem. Biodivers.* 14, e1700057. doi:10.1002/cbdv.201700057.
- 2481 Liu, N., Song, F., Shang, F., and Huang, Y. (2015). Mycemycins A–E, New

- 2482 Dibenzoazepinones Isolated from Two Different Streptomyces. *Mar. Drugs*  
2483 13, 6247–6258. doi:10.3390/md13106247.
- 2484 Liu, Q., Xiao, L., Zhou, Y., Deng, K., Tan, G., Han, Y., et al. (2016). Development of  
2485 *Streptomyces* sp. FR-008 as an emerging chassis. *Synth. Syst. Biotechnol.* 1,  
2486 207–214. doi:10.1016/J.SYNBIO.2016.07.002.
- 2487 Losada, A. A., Cano-Prieto, C., García-Salcedo, R., Braña, A. F., Méndez, C., Salas, J.  
2488 A., et al. (2017). Caboxamycin biosynthesis pathway and identification of novel  
2489 benzoxazoles produced by cross-talk in *Streptomyces* sp. NTK 937. *Microb.*  
2490 *Biotechnol.* 10, 873–885. doi:10.1111/1751-7915.12716.
- 2491 Luo, Y., Huang, H., Liang, J., Wang, M., Lu, L., Shao, Z., et al. (2013). Activation and  
2492 characterization of a cryptic polycyclic tetramate macrolactam biosynthetic  
2493 gene cluster. *Nat. Commun.* 4, 2894. doi:10.1038/ncomms3894.
- 2494 MacMillan, S. V., Alexander, D. A., Culham, D. E., Kunte, H. J., Marshall, E. V.,  
2495 Rochon, D., et al. (1999). The ion coupling and organic substrate specificities of  
2496 osmoregulatory transporter ProP in *Escherichia coli*. *Biochim. Biophys. Acta -*  
2497 *Biomembr.* 1420, 30–44. doi:10.1016/S0005-2736(99)00085-1.
- 2498 Maddocks, S. E., and Oyston, P. C. F. (2008). Structure and function of the LysR-  
2499 type transcriptional regulator (LTTR) family proteins. *Microbiology* 154, 3609–  
2500 3623. doi:10.1099/mic.0.2008/022772-0.
- 2501 Magarvey, N. A., Haltli, B., He, M., Greenstein, M., and Hucul, J. A. (2006).  
2502 Biosynthetic Pathway for Mannopeptimycins, Lipoglycopeptide Antibiotics  
2503 Active against Drug-Resistant Gram-Positive Pathogens. *Antimicrob. Agents*  
2504 *Chemother.* 50, 2167–2177. doi:ICN33.
- 2505 Majiduddin, F. K., Materon, I. C., and Palzkill, T. G. (2002). Molecular analysis of  
2506 beta-lactamase structure and function. *Int. J. Med. Microbiol.* 292, 127–137.  
2507 doi:10.1078/1438-4221-00198.
- 2508 Manivasagan, P., Kang, K.-H., Sivakumar, K., Li-Chan, E. C. Y., Oh, H.-M., and Kim,  
2509 S.-K. (2014). Marine actinobacteria: An important source of bioactive natural  
2510 products. *Environ. Toxicol. Pharmacol.* 38, 172–188.  
2511 doi:10.1016/j.etap.2014.05.014.
- 2512 Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L.  
2513 Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.*  
2514 43, D222–D226. doi:10.1093/nar/gku1221.
- 2515 Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-  
2516 Scott, C., et al. (2011). CDD: a Conserved Domain Database for the functional  
2517 annotation of proteins. *Nucleic Acids Res.* 39, D225–D229.

- doi:10.1093/nar/gkq1189.
- Mascher, T., Zimmer, S. L., Smith, T.-A., and Helmann, J. D. (2004). Antibiotic-Inducible Promoter Regulated by the Cell Envelope Stress-Sensing Two-Component System LiaRS of *Bacillus subtilis*. *Antimicrob. Agents Chemother.* 48, 2888–2896. doi:10.1128/AAC.48.8.2888-2896.2004.
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification. *Nat. Chem. Biol.* 11, 625–631. doi:10.1038/nchembio.1890.
- Mehbub, M. F., Tanner, J. E., Barnett, S. J., Franco, C. M. M., and Zhang, W. (2016). The role of sponge-bacteria interactions: the sponge *Aplysilla rosea* challenged by its associated bacterium *Streptomyces* ACT-52A in a controlled aquarium system. *Appl. Microbiol. Biotechnol.* 100, 10609–10626. doi:10.1007/s00253-016-7878-9.
- Miller, K. J., and Wood, J. M. (1996). Osmoadaptation by rhizosphere bacteria. *Annu. Rev. Microbiol.* 50, 101–136. doi:10.1146/annurev.micro.50.1.101.
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2018). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* doi:10.1093/nar/gky1100.
- Molina-Henares, A. J., Krell, T., Eugenia Guazzaroni, M., Segura, A., and Ramos, J. L. (2006). Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol. Rev.* 30, 157–186. doi:10.1111/j.1574-6976.2005.00008.x.
- Montalbán-López, M., Deng, J., van Heel, A. J., and Kuipers, O. P. (2018). Specificity and Application of the Lantibiotic Protease NisP. *Front. Microbiol.* 9, 160. doi:10.3389/fmicb.2018.00160.
- Motohashi, K., Takagi, M., Yamamura, H., Hayakawa, M., and Shin-ya, K. (2010). A new angucycline and a new butenolide isolated from lichen-derived *Streptomyces* spp. *J. Antibiot. (Tokyo)*. 63, 545–548. doi:10.1038/ja.2010.94.
- Naas, T., Oueslati, S., Bonnin, R. A., Dabos, M. L., Zavala, A., Dortet, L., et al. (2017). Beta-lactamase database (BLDB) – structure and function. *J. Enzyme Inhib. Med. Chem.* 32, 917–919. doi:10.1080/14756366.2017.1344235.
- Nara, A., Hashimoto, T., Komatsu, M., Nishiyama, M., Kuzuyama, T., and Ikeda, H. (2017). Characterization of bafilomycin biosynthesis in *Kitasatospora setae* KM-6054 and comparative analysis of gene clusters in Actinomycetales microorganisms. *J. Antibiot. (Tokyo)*. 70, 616–624. doi:10.1038/ja.2017.33.
- Navarro-Muñoz, J. C., Selem-Mojica, N., Mullaney, M. W., Kautsar, S., Tryon, J. H.,

- 2554 Parkinson, E. I., et al. (2018). A computational framework for systematic  
2555 exploration of biosynthetic diversity from large-scale genomic data. *BioRxiv*,  
2556 445270. <https://doi.org/10.1101/445270>.
- 2557 Nguyen, T. B., Kitani, S., Shimma, S., and Nihira, T. (2018). Butenolides from  
2558 *Streptomyces albus* J1074 Act as External Signals To Stimulate Avermectin  
2559 Production in *Streptomyces avermitilis*. *Appl. Environ. Microbiol.* 84,  
2560 AEM.02791-17. doi:10.1128/AEM.02791-17.
- 2561 Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., et al.  
2562 (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an  
2563 Unrelated Decapeptide, Surugamide F. *ChemBioChem* 17, 1709–1712.  
2564 doi:10.1002/cbic.201600350.
- 2565 Nishida, H. (2012). Evolution of genome base composition and genome size in  
2566 bacteria. *Front. Microbiol.* 3, 420. doi:10.3389/fmicb.2012.00420.
- 2567 Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et  
2568 al. (2015). Roary: rapid large-scale prokaryote pan genome analysis.  
2569 *Bioinformatics* 31, 3691–3693. doi:10.1093/bioinformatics/btv421.
- 2570 Paulus, C., Rebets, Y., Tokovenko, B., Nadmid, S., Terekhova, L. P., Myronovskiy,  
2571 M., et al. (2017). New natural products identified by combined genomics-  
2572 metabolomics profiling of marine *Streptomyces* sp. MP131-18. *Sci. Rep.* 7,  
2573 42382. doi:10.1038/srep42382.
- 2574 Penn, K., and Jensen, P. R. (2012). Comparative genomics reveals evidence of  
2575 marine adaptation in *Salinispora* species. *BMC Genomics* 13, 86.  
2576 doi:10.1186/1471-2164-13-86.
- 2577 Qurashi, A. W., and Sabri, A. N. (2011). Osmoadaptation and plant growth  
2578 promotion by salt tolerant bacteria under salt stress. *African J. Microbiol. Res.*  
2579 5, 3546–3554. doi:10.5897/AJMR11.736.
- 2580 R Core Team (2018). R: A Language and Environment for Statistical Computing.  
2581 Available at: <https://www.r-project.org>.
- 2582 Racine, J. S. (2012). RSTUDIO: A platform-independent IDE for R and sweave. *J.*  
2583 *Appl. Econom.* 27, 167–172. doi:10.1002/jae.1278.
- 2584 Rademacher, C., and Masepohl, B. (2012). Copper-responsive gene regulation in  
2585 bacteria. *Microbiology* 158, 2451–2464. doi:10.1099/mic.0.058487-0.
- 2586 Rashad, F. M., Fathy, H. M., El-Zayat, A. S., and Elghonaimy, A. M. (2015). Isolation  
2587 and characterization of multifunctional *Streptomyces* species with  
2588 antimicrobial, nematicidal and phytohormone activities from marine  
2589 environments in Egypt. *Microbiol. Res.* 175, 34–47.

- 2590 doi:10.1016/J.MICRES.2015.03.002.
- 2591 Rasko, D. A., Rosovitz, M. J., Økstad, O. A., Fouts, D. E., Jiang, L., Cer, R. Z., et al.  
 2592 (2007). Complete sequence analysis of novel plasmids from emetic and  
 2593 periodontal *Bacillus cereus* isolates reveals a common evolutionary history  
 2594 among the *B. cereus*-group plasmids, including *Bacillus anthracis* pXO1. *J.*  
 2595 *Bacteriol.* 189, 52–64. doi:10.1128/JB.01313-06.
- 2596 Reimer, A., Blohm, A., Quack, T., Greveling, C. G., Kozjak-Pavlovic, V., Rudel, T.,  
 2597 et al. (2015). Inhibitory activities of the marine streptomycete-derived  
 2598 compound SF2446A2 against *Chlamydia trachomatis* and *Schistosoma*  
 2599 *mansoni*. *J. Antibiot. (Tokyo)*. 68, 674–679. doi:10.1038/ja.2015.54.
- 2600 Reynolds, D., and Thomas, T. (2016). Evolution and function of eukaryotic-like  
 2601 proteins from sponge symbionts. *Mol. Ecol.* 25, 5242–5253.  
 2602 doi:10.1111/mec.13812.
- 2603 Rigali, S., Anderssen, S., Naômé, A., and van Wezel, G. P. (2018). Cracking the  
 2604 regulatory code of biosynthetic gene clusters as a strategy for natural product  
 2605 discovery. *Biochem. Pharmacol.* 153, 24–34. doi:10.1016/j.bcp.2018.01.007.
- 2606 Robbel, L., and Marahiel, M. A. (2010). Daptomycin, a bacterial lipopeptide  
 2607 synthesized by a nonribosomal machinery. *J. Biol. Chem.* 285, 27501–8.  
 2608 doi:10.1074/jbc.R110.128181.
- 2609 Rodríguez, H., Rico, S., Díaz, M., and Santamaría, R. I. (2013). Two-component  
 2610 systems in *Streptomyces*: key regulators of antibiotic complex pathways.  
 2611 *Microb. Cell Fact.* 12, 127. doi:10.1186/1475-2859-12-127.
- 2612 Roessler, M., and Muller, V. (2001). Osmoadaptation in bacteria and archaea:  
 2613 common principles and differences. *Environ. Microbiol.* 3, 743–754.  
 2614 doi:10.1046/j.1462-2920.2001.00252.x.
- 2615 Rolain, J.-M., Abat, C., Jimeno, M.-T., Fournier, P.-E., and Raoult, D. (2016). Do we  
 2616 need new antibiotics? *Clin. Microbiol. Infect.* 22, 408–415.  
 2617 doi:10.1016/j.cmi.2016.03.012.
- 2618 Romero, D., Traxler, M. F., López, D., and Kolter, R. (2011). Antibiotics as Signal  
 2619 Molecules. *Chem. Rev.* 111, 5492–5505. doi:10.1021/cr2000509.
- 2620 Rong, X., Doroghazi, J. R., Cheng, K., Zhang, L., Buckley, D. H., and Huang, Y.  
 2621 (2013). Classification of *Streptomyces* phylogroup *pratensis* (Doroghazi and  
 2622 Buckley, 2010) based on genetic and phenotypic evidence, and proposal of  
 2623 *Streptomyces pratensis* sp. nov. *Syst. Appl. Microbiol.* 36, 401–407.  
 2624 doi:10.1016/j.syapm.2013.03.010.

- 2625 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et  
 2626 al. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model  
 2627 Choice Across a Large Model Space. *Syst. Biol.* 61, 539–542.  
 2628 doi:10.1093/sysbio/sys029.
- 2629 RStudio Team (2015). RStudio: Integrated Development Environment for R.  
 2630 Available at: <http://www.rstudio.com/>.
- 2631 Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al.  
 2632 (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–  
 2633 945. doi:10.1093/bioinformatics/16.10.944.
- 2634 Rutledge, P. J., and Challis, G. L. (2015). Discovery of microbial natural products by  
 2635 activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* 13, 509–523.  
 2636 doi:10.1038/nrmicro3496.
- 2637 Salwan, R., and Sharma, V. (2018). “The Role of Actinobacteria in the Production of  
 2638 Industrial Enzymes,” in *New and Future Developments in Microbial  
 2639 Biotechnology and Bioengineering* (Elsevier), 165–177. doi:10.1016/B978-0-444-  
 2640 63994-3.00011-4.
- 2641 Schmid, M., Muri, J., Melidis, D., Varadarajan, A. R., Somerville, V., Wicki, A., et al.  
 2642 (2018). Comparative Genomics of Completely Sequenced *Lactobacillus*  
 2643 *helveticus* Genomes Provides Insights into Strain-Specific Genes and Resolves  
 2644 Metagenomics Data Down to the Strain Level. *Front. Microbiol.* 9, 63.  
 2645 doi:10.3389/fmicb.2018.00063.
- 2646 Schneemann, I., Kajahn, I., Ohlendorf, B., Zinecker, H., Erhard, A., Nagel, K., et al.  
 2647 (2010). Mayamycin, a Cytotoxic Polyketide from a *Streptomyces* Strain Isolated  
 2648 from the Marine Sponge *Halichondria panicea*. *J. Nat. Prod.* 73, 1309–1312.  
 2649 doi:10.1021/np100135b.
- 2650 Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*  
 2651 30, 2068–2069. doi:10.1093/bioinformatics/btu153.
- 2652 Seipke, R. F., Barke, J., Brearley, C., Hill, L., Yu, D. W., Goss, R. J. M., et al. (2011). A  
 2653 Single *Streptomyces* Symbiont Makes Multiple Antifungals to Support the  
 2654 Fungus Farming Ant *Acromyrmex octospinosus*. *PLoS One* 6, e22028.  
 2655 doi:10.1371/journal.pone.0022028.
- 2656 Seipke, R. F., Kaltenpoth, M., and Hutchings, M. I. (2012). *Streptomyces* as  
 2657 symbionts: an emerging and widespread theme? *FEMS Microbiol. Rev.* 36,  
 2658 862–876. doi:10.1111/j.1574-6976.2011.00313.x.
- 2659 Ser, H.-L., Tan, L. T.-H., Law, J. W.-F., Chan, K.-G., Duangjai, A., Saokaew, S., et al.  
 2660 (2017). Focused Review: Cytotoxic and Antioxidant Potentials of Mangrove-

- 2661        Derived *Streptomyces*. *Front. Microbiol.* 8, 2065. doi:10.3389/fmicb.2017.02065.
- 2662        Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al.  
2663        (2003). Cytoscape: a software environment for integrated models of  
2664        biomolecular interaction networks. *Genome Res.* 13, 2498–504.  
2665        doi:10.1101/gr.1239303.
- 2666        Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide  
2667        synthase paradigms. *Curr. Opin. Chem. Biol.* 7, 285–295. doi:10.1016/S1367-  
2668        5931(03)00020-6.
- 2669        Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011). Production of  
2670        Lantipeptides in *Escherichia coli*. *J. Am. Chem. Soc.* 133, 2338–2341.  
2671        doi:10.1021/ja109044r.
- 2672        Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T.-J., Lee, J. E., et al. (2013).  
2673        Advantages of Single-Molecule Real-Time Sequencing in High-GC Content  
2674        Genomes. *PLoS One* 8, e68824. doi:10.1371/journal.pone.0068824.
- 2675        Singh, M. P., Petersen, P. J., Weiss, W. J., Janso, J. E., Luckman, S. W., Lenoy, E. B., et  
2676        al. (2003). Mannopeptimycins, New Cyclic Glycopeptide Antibiotics Produced  
2677        by *Streptomyces hygroscopicus* LL-AC98: Antibacterial and Mechanistic  
2678        Activities. *Antimicrob. Agents Chemother.* 47, 62–69. doi:10.1128/AAC.47.1.62-  
2679        69.2003.
- 2680        Slaby, B. M., Hackl, T., Horn, H., Bayer, K., and Hentschel, U. (2017). Metagenomic  
2681        binning of a marine sponge microbiome reveals unity in defense but metabolic  
2682        specialization. *ISME J.* 11, 2465–2478. doi:10.1038/ismej.2017.101.
- 2683        Song, F., Liu, N., Liu, M., Chen, Y., and Huang, Y. (2018). Identification and  
2684        Characterization of Mycemycin Biosynthetic Gene Clusters in *Streptomyces*  
2685        *olivaceus* FXJ8.012 and *Streptomyces* sp. FXJ1.235. *Mar. Drugs* 16, 98.  
2686        doi:10.3390/md16030098.
- 2687        Stock, J. B., Stock, A. M., and Mottonen, J. M. (1990). Signal transduction in bacteria.  
2688        *Nature* 344, 395–400. doi:10.1038/344395a0.
- 2689        Sun, W., Liu, C., Zhang, F., Zhao, M., and Li, Z. (2018). Comparative genomics  
2690        provides insights into the marine adaptation in sponge-derived *Kocuria flava*  
2691        S43. *Front. Microbiol.* 9, 1257. doi:10.3389/fmicb.2018.01257.
- 2692        Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky,  
2693        L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids*  
2694        *Res.* 44, 6614–6624. doi:10.1093/nar/gkw569.
- 2695        Teichmann, L., Kümmel, H., Warmbold, B., and Bremer, E. (2018). OpuF, a New



- 2696 *Bacillus* Compatible Solute ABC Transporter with a Substrate-Binding Protein  
 2697 Fused to the Transmembrane Domain. *Appl. Environ. Microbiol.* 84, e01728-18.  
 2698 doi:10.1128/AEM.01728-18.
- 2699 ter Beek, J., Guskov, A., and Slotboom, D. J. (2014). Structural diversity of ABC  
 2700 transporters. *J. Gen. Physiol.* 143, 419–435. doi:10.1085/jgp.201411164.
- 2701 Tetzlaff, C. N., You, Z., Cane, D. E., Takamatsu, S., Omura, S., and Ikeda, H. (2006).  
 2702 A gene cluster for biosynthesis of the sesquiterpenoid antibiotic  
 2703 pentalenolactone in *Streptomyces avermitilis*. *Biochemistry* 45, 6179–6186.  
 2704 doi:10.1021/bi060419n.
- 2705 Thabit, A. K., Crandon, J. L., and Nicolau, D. P. (2015). Antimicrobial resistance:  
 2706 impact on clinical and economic outcomes and the need for new  
 2707 antimicrobials. *Expert Opin. Pharmacother.* 16, 159–177.  
 2708 doi:10.1517/14656566.2015.993381.
- 2709 Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., et al.  
 2710 (2010). Functional genomic signatures of sponge bacteria reveal unique and  
 2711 shared features of symbiosis. *ISME J.* 4, 1557–1567. doi:10.1038/ismej.2010.74.
- 2712 Tian, R.-M., Wang, Y., Bougouffa, S., Gao, Z.-M., Cai, L., Zhang, W.-P., et al. (2014).  
 2713 Effect of Copper Treatment on the Composition and Function of the Bacterial  
 2714 Community in the Sponge *Haliclona cymaeformis*. *mBio* 5, e01980.  
 2715 doi:10.1128/mBio.01980-14.
- 2716 Tian, X., Zhang, Z., Yang, T., Chen, M., Li, J., Chen, F., et al. (2016). Comparative  
 2717 Genomics Analysis of *Streptomyces* Species Reveals Their Adaptation to the  
 2718 Marine Environment and Their Diversity at the Genomic Level. *Front.*  
 2719 *Microbiol.* 7, 998. doi:10.3389/fmicb.2016.00998.
- 2720 Tierrafría, V. H., Ramos-Aboites, H. E., Gosset, G., and Barona-Gómez, F. (2011).  
 2721 Disruption of the siderophore-binding desE receptor gene in *Streptomyces*  
 2722 *coelicolor* A3(2) results in impaired growth in spite of multiple iron-  
 2723 siderophore transport systems. *Microb. Biotechnol.* 4, 275–85.  
 2724 doi:10.1111/j.1751-7915.2010.00240.x.
- 2725 UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids*  
 2726 *Res.* 43, D204–D212. doi:10.1093/nar/gku989.
- 2727 van der Meij, A., Worsley, S. F., Hutchings, M. I., and van Wezel, G. P. (2017).  
 2728 Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol.*  
 2729 *Rev.* 41, 392–416. doi:10.1093/femsre/fux005.
- 2730 Van Melderren, L. (2010). Toxin–antitoxin systems: why so many, what for? *Curr.*  
 2731 *Opin. Microbiol.* 13, 781–785. doi:10.1016/j.mib.2010.10.006.

- 2732 Vicente, C., Thibessard, A., Lorenzi, J.-N., Benhadj, M., Hôtel, L., Gacemi-Kirane, D.,  
 2733 et al. (2018). Comparative Genomics among Closely Related *Streptomyces*  
 2734 Strains Revealed Specialized Metabolite Biosynthetic Gene Cluster Diversity.  
 2735 *Antibiotics* 7, 86. doi:10.3390/antibiotics7040086.
- 2736 Volff, J. N., Viell, P., and Altenbuchner, J. (1997). Artificial circularization of the  
 2737 chromosome with concomitant deletion of its terminal inverted repeats  
 2738 enhances genetic instability and genome rearrangement in *Streptomyces*  
 2739 *lividans*. *Mol. Gen. Genet.* 253, 753–760. doi:10.1007/s004380050380.
- 2740 Waddell, P. J., and Steel, M. . (1997). General Time-Reversible Distances with  
 2741 Unequal Rates across Sites: Mixing  $\Gamma$  and Inverse Gaussian Distributions with  
 2742 Invariant Sites. *Mol. Phylogenet. Evol.* 8, 398–414. doi:10.1006/mpev.1997.0452.
- 2743 Wang, T., Chen, Z., Cheng, Q., Zhou, M., Tian, X., Xie, P., et al. (2012).  
 2744 Characterization of replication and conjugation of plasmid pWTY27 from a  
 2745 widely distributed *Streptomyces* species. *BMC Microbiol.* 12, 253.  
 2746 doi:10.1186/1471-2180-12-253.
- 2747 Weaver, D., Karoonuthaisiri, N., Tsai, H.-H., Huang, C.-H., Ho, M.-L., Gai, S., et al.  
 2748 (2004). Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S.*  
 2749 *coelicolor* A3(2) chromosome. *Mol. Microbiol.* 51, 1535–1550.  
 2750 doi:10.1111/j.1365-2958.2003.03920.x.
- 2751 Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat.*  
 2752 *Softw.* 40, 1–29. doi:10.18637/jss.v040.i01.
- 2753 Wilkens, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Rep.*  
 2754 7, 14. doi:10.12703/P7-14.
- 2755 Wilson, K. (2001). “Preparation of Genomic DNA from Bacteria,” in *Current*  
 2756 *Protocols in Molecular Biology* (Hoboken, NJ, USA: John Wiley & Sons, Inc.),  
 2757 Unit 2.4. doi:10.1002/0471142727.mb0204s56.
- 2758 Wood, J. M. (2007). “Bacterial Osmosensing Transporters,” in *Methods in*  
 2759 *Enzymology* (Academic Press), 77–107. doi:10.1016/S0076-6879(07)28005-X.
- 2760 Wood, J. M. (2015). Bacterial responses to osmotic challenges. *J. Gen. Physiol.* 145,  
 2761 381–388. doi:10.1085/jgp.201411296.
- 2762 Wright, G. D. (2005). Bacterial resistance to antibiotics: Enzymatic degradation and  
 2763 modification. *Adv. Drug Deliv. Rev.* 57, 1451–1470.  
 2764 doi:10.1016/j.addr.2005.04.002.
- 2765 Wright, G. D. (2012). “The Origins of Antibiotic Resistance,” in *Antibiotic*  
 2766 *Resistance* (Springer, Berlin, Heidelberg), 13–30. doi:10.1007/978-3-642-28951-

2767 4\_2.

- 2768 Wu, H., Qu, S., Lu, C., Zheng, H., Zhou, X., Bai, L., et al. (2012). Genomic and  
2769 transcriptomic insights into the thermo-regulated biosynthesis of validamycin  
2770 in *Streptomyces hygroscopicus* 5008. *BMC Genomics* 13, 337. doi:10.1186/1471-  
2771 2164-13-337.
- 2772 Wyche, T. P., Ruzzini, A. C., Schwab, L., Currie, C. R., and Clardy, J. (2017).  
2773 Tryptorubin A: A Polycyclic Peptide from a Fungus-Derived Streptomycete. *J.*  
2774 *Am. Chem. Soc.* 139, 12899–12902. doi:10.1021/jacs.7b06176.
- 2775 Xie, C.-L., Xia, J.-M., Wang, J.-S., Lin, D.-H., and Yang, X.-W. (2018). Metabolomic  
2776 Investigations on *Nesterenkonia flava* Revealed Significant Differences  
2777 between Marine and Terrestrial Actinomycetes. *Mar. Drugs* 16, 356.  
2778 doi:10.3390/md16100356.
- 2779 Xu, X.-N., Chen, L.-Y., Chen, C., Tang, Y.-J., Bai, F.-W., Su, C., et al. (2018). Genome  
2780 Mining of the Marine Actinomycete *Streptomyces* sp. DUT11 and Discovery of  
2781 Tunicamycins as Anti-complement Agents. *Front. Microbiol.* 9, 1318.  
2782 doi:10.3389/fmicb.2018.01318.
- 2783 Xu, Y., Guo, J., Wang, L., Jiang, R., Jin, X., Liu, J., et al. (2016). The Crystal Structure  
2784 of the YknZ Extracellular Domain of ABC Transporter YknWXYZ from  
2785 *Bacillus amyloliquefaciens*. *PLoS One* 11, e0155846.  
2786 doi:10.1371/journal.pone.0155846.
- 2787 Yaakop, A. S., Chan, K.-G., Ee, R., Lim, Y. L., Lee, S.-K., Manan, F. A., et al. (2016).  
2788 Characterization of the mechanism of prolonged adaptation to osmotic stress  
2789 of *Jeotgalibacillus malaysiensis* via genome and transcriptome sequencing  
2790 analyses. *Sci. Rep.* 6, 33660. doi:10.1038/srep33660.
- 2791 Yagüe, P., Lopez-Garcia, M. T., Rioseras, B., Sanchez, J., and Manteca, A. (2012).  
2792 New insights on the development of *Streptomyces* and their relationships with  
2793 secondary metabolite production. *Curr. trends Microbiol.* 8, 65–73.
- 2794 Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., et al.  
2795 (2015). Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad.*  
2796 *Sci.* 112, 857–862. doi:10.1073/pnas.1422108112.
- 2797 Yuwono, T. (2005). Metabolism of betaine as a carbon source by an osmotolerant  
2798 bacterium isolated from the weed rhizosphere. *World J. Microbiol. Biotechnol.*  
2799 21, 69–73. doi:10.1007/s11274-004-1935-8.
- 2800 Zhang, C., Yang, Z., Qin, X., Ma, J., Sun, C., Huang, H., et al. (2018). Genome  
2801 Mining for Mycemycin: Discovery and Elucidation of Related Methylation and  
2802 Chlorination Biosynthetic Chemistries. *Org. Lett.* 20, 7633–7636.

- 2803       doi:10.1021/acs.orglett.8b03373.
- 2804       Zhang, W., Fortman, J. L., Carlson, J. C., Yan, J., Liu, Y., Bai, F., et al. (2013).  
 2805       Characterization of the Bafilomycin Biosynthetic Gene Cluster from  
 2806       *Streptomyces lohii*. *ChemBioChem* 14, 301–306. doi:10.1002/cbic.201200743.
- 2807       Zhang, X.-Y., Han, X.-X., Chen, X.-L., Dang, H.-Y., Xie, B.-B., Qin, Q.-L., et al. (2015).  
 2808       Diversity of cultivable protease-producing bacteria in sediments of Jiaozhou  
 2809       Bay, China. *Front. Microbiol.* 6, 1021. doi:10.3389/fmicb.2015.01021.
- 2810       Zhang, Z., Pan, H.-X., and Tang, G.-L. (2017). New insights into bacterial type II  
 2811       polyketide biosynthesis. *F1000Research* 6, 172.  
 2812       doi:10.12688/f1000research.10466.1.
- 2813       Zhou, Z., Gu, J., Li, Y.-Q., and Wang, Y. (2012). Genome plasticity and systems  
 2814       evolution in *Streptomyces*. *BMC Bioinformatics* 13, S8. doi:10.1186/1471-2105-  
 2815       13-S10-S8.
- 2816       Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in  
 2817       microbes – a review. *Nat. Prod. Rep.* 33, 988–1005. doi:10.1039/C6NP00025H.

2818 7. Supplementary material

Cluster	Type	From	To	Most similar known cluster
Cluster 1	Lantipeptide-T1pks-Nrps	9520	282674	Candididin biosynthetic gene cluster (100% of genes show similarity)
Cluster 2	T3pks	287634	328731	Herboxidiene biosynthetic gene cluster (12% of genes show similarity)
Cluster 3	Bacteriocin-Terpene	454168	486617	Carotenoid biosynthetic gene cluster (54% of genes show similarity)
Cluster 4	Ectoine	1226907	1237305	Ectoine biosynthetic gene cluster (100% of genes show similarity)
Cluster 5	Siderophore	2136943	2148763	Desferrioxamine B biosynthetic gene cluster (100% of genes show similarity)
Cluster 6	Nrps	2377319	2421668	-
Cluster 7	Nrps	2881828	2987506	Mannopectimycin biosynthetic gene cluster (7% of genes show similarity)
Cluster 8	Nrps	3238171	3288460	Scabichelin biosynthetic gene cluster (40% of genes show similarity)
Cluster 9	Nrps	4012785	4073245	Mannopectimycin biosynthetic gene cluster (51% of genes show similarity)
Cluster 10	Lantipeptide	4096267	4118852	SAL-2242_biosynthetic_gene_cluster (100% of genes show similarity)
Cluster 11	Thiopeptide	4422066	4454538	-
Cluster 12	Terpene	4948699	4969673	Albaflavenone biosynthetic gene cluster (100% of genes show similarity)
Cluster 13	Terpene	5280779	5303064	Kanamycin biosynthetic gene cluster (1% of genes show similarity)
Cluster 14	Siderophore	5558173	5573204	-
Cluster 15	Nrps	5648771	5708652	Tetronasin biosynthetic gene cluster (9% of genes show similarity)
Cluster 16	Bacteriocin	5946086	5957414	-
Cluster 17	Bacteriocin	6364321	6374536	-
Cluster 18	Terpene	6440077	6466641	Hopene biosynthetic gene cluster (76% of genes show similarity)
Cluster 19	T1pks-Nrps	6501121	6550531	SGR PTMs biosynthetic gene cluster (100% of genes show similarity)
Cluster 20	Terpene-Nrps	6690082	6777548	Lividomycin biosynthetic gene cluster (10% of genes show similarity)

2819 **Table S1:** Putative smBGCs predicted to be present in the SM17 genome using the  
2820 antiSMASH program.

Cluster	Type	From	To	Most similar known cluster
Cluster 1	Terpene	29172	50203	-
Cluster 2	T2pks-Terpene	68208	117510	Spore pigment biosynthetic gene cluster (75% of genes show similarity)
Cluster 3	Melanin	151502	161990	Melanin biosynthetic gene cluster (100% of genes show similarity)
Cluster 4	T1pks	390670	438115	Surfactin biosynthetic gene cluster (8% of genes show similarity)
Cluster 5	Nrps	457604	520442	Daptomycin biosynthetic gene cluster (9% of genes show similarity)
Cluster 6	Lantipeptide	596609	619990	-
Cluster 7	T1pks-Nrps	728498	820279	Chivosazole biosynthetic gene cluster (22% of genes show similarity)
Cluster 8	Terpene	902952	924010	Steffimycin biosynthetic gene cluster (19% of genes show similarity)
Cluster 9	Ectoine	1363227	1373625	Ectoine biosynthetic gene cluster (100% of genes show similarity)
Cluster 10	T1pks-Otherks	2295950	2349406	-
Cluster 11	Lantipeptide-Lasso peptide	2379602	2423090	-
Cluster 12	Butyrolactone	3482738	3493514	Griseoviridin/viridogrisein biosynthetic gene cluster (8% of genes show similarity)
Cluster 13	T1pks	4955466	5003964	Arginomycin biosynthetic gene cluster (20% of genes show similarity)
Cluster 14	Bacteriocin	5043642	5055585	-
Cluster 15	Terpene	5125104	5146111	Carbapenem MM 4550 biosynthetic gene cluster (10% of genes show similarity)
Cluster 16	Siderophore	5488861	5503466	-
Cluster 17	Otherks-Nrps	5660928	5786416	A33853 biosynthetic gene cluster (43% of genes show similarity)
Cluster 18	Thiopeptide-Terpene	5857201	5913709	Isorenieratene biosynthetic gene cluster (100% of genes show similarity)
Cluster 19	T2pks	5974917	6017450	Pristinamycin biosynthetic gene cluster (21% of genes show similarity)
Cluster 20	Bacteriocin	6097616	6109004	-
Cluster 21	Nrps	6308570	6362642	Coelibactin biosynthetic gene cluster (100% of genes show similarity)
Cluster 22	Terpene	6757753	6784324	Hopene biosynthetic gene cluster (69% of genes show similarity)

Cluster 23	Nrps	6978302	7029194	Coelichelin biosynthetic gene cluster (100% of genes show similarity)
Cluster 24	T1pks-Nrps	7030460	7111803	Herboxidiene biosynthetic gene cluster (16% of genes show similarity)
Cluster 25	T1pks	7248399	7346791	Bafilomycin biosynthetic gene cluster (100% of genes show similarity)
Cluster 26	Nrps	7527142	7577107	Mirubactin biosynthetic gene cluster (50% of genes show similarity)

---

2821 **Table S2:** Putative smBGCs predicted to be present in the SM18  
2822 genome using the antiSMASH program.

Cluster	Type	From	To	Most similar known cluster
Cluster 1	T1pks-Nrps	3011	61711	-
Cluster 2	T1pks-Nrps	224752	274162	SGR PTMs biosynthetic gene cluster (100% of genes show similarity)
Cluster 3	Terpene	308626	335190	Hopene biosynthetic gene cluster (76% of genes show similarity)
Cluster 4	Bacteriocin	415688	425903	-
Cluster 5	Otherks	669953	711002	Avermectin biosynthetic gene cluster (66% of genes show similarity)
Cluster 6	Bacteriocin	879961	891289	-
Cluster 7	Nrps	1136316	1199422	Tetronasin biosynthetic gene cluster (9% of genes show similarity)
Cluster 8	Siderophore	1268164	1283196	-
Cluster 9	Terpene	1531774	1554059	Kanamycin biosynthetic gene cluster (1% of genes show similarity)
Cluster 10	Terpene	1865241	1886215	Albaflavenone biosynthetic gene cluster (100% of genes show similarity)
Cluster 11	Thiopeptide	2376688	2409159	-
Cluster 12	Bacteriocin	2560714	2571226	Goadsporin biosynthetic gene cluster (12% of genes show similarity)
Cluster 13	Lantipeptide	2694944	2735867	SAL-2242 biosynthetic gene cluster (100% of genes show similarity)
Cluster 14	Nrps	3553726	3604015	Scabichelin biosynthetic gene cluster (40% of genes show similarity)
Cluster 15	Nrps	3877105	3982777	Mannopectimycin biosynthetic gene cluster (7% of genes show similarity)
Cluster 16	Nrps	4469477	4513826	-
Cluster 17	Siderophore	4740450	4752270	Desferrioxamine B biosynthetic gene cluster (100% of genes show similarity)
Cluster 18	Ectoine	5635346	5645744	Ectoine biosynthetic gene cluster (100% of genes show similarity)
Cluster 19	Other	6337383	6381213	Indigoidine biosynthetic gene cluster (80% of genes show similarity)
Cluster 20	Bacteriocin-Terpene	6396161	6430146	Carotenoid biosynthetic gene cluster (54% of genes show similarity)
Cluster 21	T3pks	6520374	6561471	Herboxidiene biosynthetic gene cluster (12% of genes show similarity)
Cluster 22	Lantipeptide-T1pks-Nrps	6566423	6838639	Candidicin biosynthetic gene cluster (100% of genes show similarity)

2823 **Table S3:** Putative smBGCs predicted to be present in the *S. albus* J1074 genome using the antiSMASH program.



Cluster	Type	From	To	Most similar known cluster
Cluster 1	Blactam-T1pks-Nrps	70873	221343	Carbapenem MM 4550 biosynthetic gene cluster (65% of genes show similarity)
Cluster 2	Nrps	327693	378596	Coelichelin biosynthetic gene cluster (90% of genes show similarity)
Cluster 3	Terpene	393138	418630	Isorenieratene biosynthetic gene cluster (100% of genes show similarity)
Cluster 4	Bacteriocin	557521	568045	-
Cluster 5	Blactam	634715	658209	Clavulanic acid biosynthetic gene cluster (20% of genes show similarity)
Cluster 6	Terpene	738599	765184	Hopene biosynthetic gene cluster (69% of genes show similarity)
Cluster 7	T1pks	1104244	1198984	Vicenistatin biosynthetic gene cluster (60% of genes show similarity)
Cluster 8	Bacteriocin	1409256	1420548	-
Cluster 9	Nrps	1560425	1622443	Arylomycin biosynthetic gene cluster (22% of genes show similarity)
Cluster 10	Siderophore	1894420	1909004	-
Cluster 11	Terpene	1973589	1994659	-
Cluster 12	Bacteriocin	2151745	2161966	-
Cluster 13	Butyrolactone	3819031	3829957	Lactonamycin biosynthetic gene cluster (3% of genes show similarity)
Cluster 14	T1pks-Nrps	4023172	4080102	Istamycin biosynthetic gene cluster (11% of genes show similarity)
Cluster 15	Siderophore	4727646	4739427	Desferrioxamine B biosynthetic gene cluster (83% of genes show similarity)
Cluster 16	Lantipeptide	4792335	4815403	-
Cluster 17	Terpene	5226078	5247109	-
Cluster 18	Ectoine	5705549	5715947	Ectoine biosynthetic gene cluster (100% of genes show similarity)
Cluster 19	T2pks-Otherks	6079056	6136319	Cinerubin B biosynthetic gene cluster (28% of genes show similarity)
Cluster 20	Terpene	6181067	6202113	Steffimycin biosynthetic gene cluster (19% of genes show similarity)
Cluster 21	Ectoine-Terpene	6482662	6509101	Ectoine biosynthetic gene cluster (100% of genes show similarity)
Cluster 22	Bacteriocin	6532212	6542439	-
Cluster 23	T3pks	6695745	6736803	Tetronasin biosynthetic gene cluster (11% of genes show similarity)

Cluster 24	Melanin	6830784	6841248	Melanin biosynthetic gene cluster (100% of genes show similarity)
Cluster 25	T2pks-Terpene	6882032	6931236	Spore pigment biosynthetic gene cluster (83% of genes show similarity)
Cluster 26	Nrps	7082613	7132319	Zorbamycin biosynthetic gene cluster (6% of genes show similarity)
Cluster 27	Butyrolactone	7211688	7222623	-

---

2824 **Table S4:** Putative smBGCs predicted to be present in the *S. pratensis*  
2825 ATCC 33331 genome using the antiSMASH program.

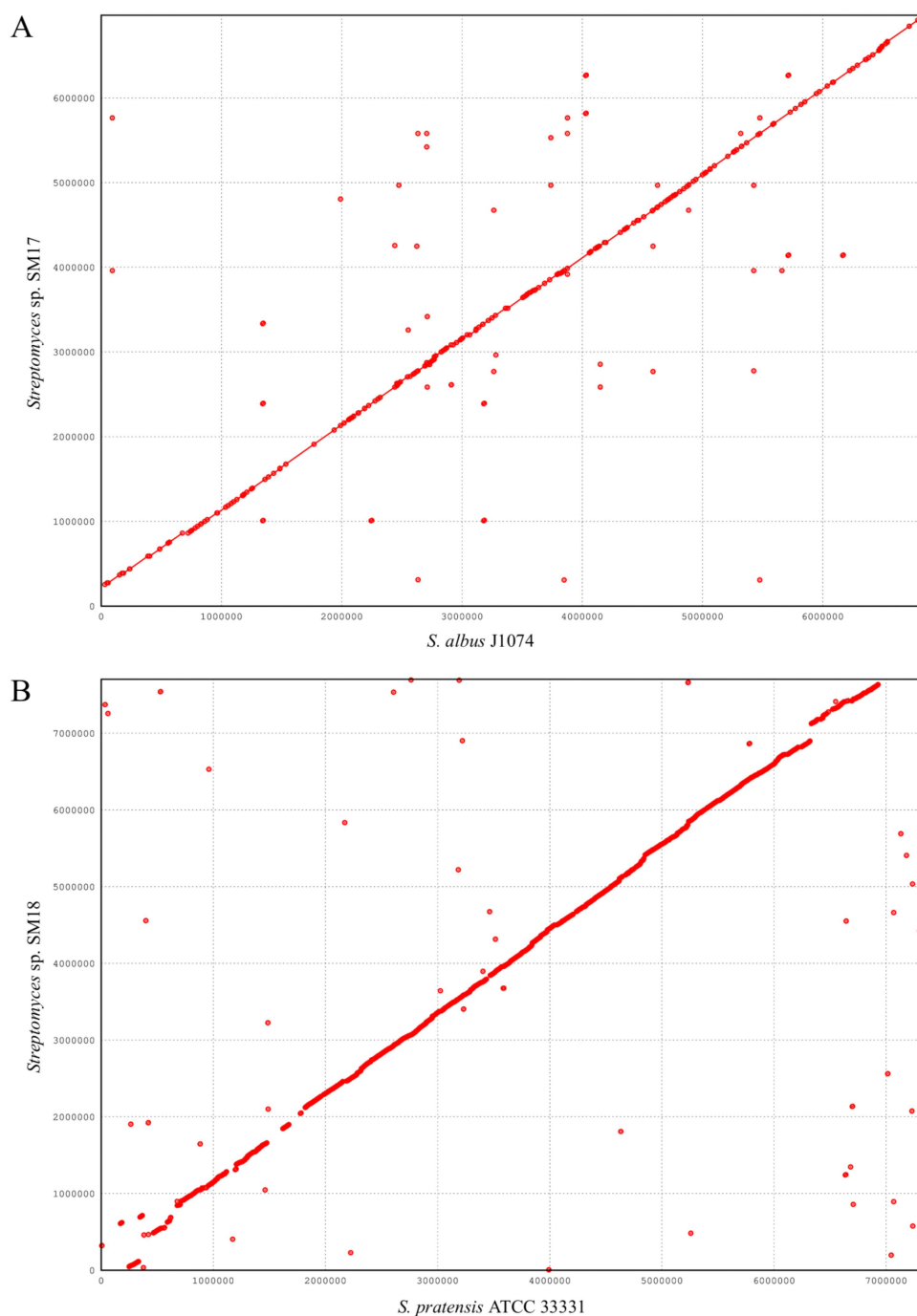
Gene	Annotations
<i>adhD</i> <sup>a</sup>	NDMA-dependent alcohol dehydrogenase / Zinc-binding alcohol dehydrogenase
<i>ahcY</i> <sup>a</sup>	Adenosylhomocysteinase
<i>aprX</i> <sup>*</sup>	Serine protease AprX / Subtilase family protein / Peptidase S8
<i>bepR</i> <sup>*</sup>	HTH-type transcriptional repressor BepR / TetR family transcriptional regulator
<i>bioC</i> <sup>b</sup>	Malonyl-[acyl-carrier protein] O-methyltransferase / Class I SAM-dependent methyltransferase
<i>cpnA</i> <sup>*</sup>	Cyclopentanol dehydrogenase / SDR family oxidoreductase
<i>cynR</i> <sup>a</sup>	HTH-type transcriptional regulator CynR / LysR family transcriptional regulator
<i>degU</i> <sup>a</sup>	Transcriptional regulatory protein DegU / DNA-binding response regulator
<i>fccA</i> <sup>*</sup>	Fumarate reductase flavoprotein subunit / FAD-dependent oxidoreductase
<i>folQ</i> <sup>*</sup>	Putative DHNTP pyrophosphohydrolase / NUDIX hydrolase
group_1044	Integrase core domain / IS3 family transposase
group_1217	Fumarylacetoacetase
group_1272	Toxin-antitoxin system, RelE family
group_1944	Restriction endonuclease
group_1945	IS3 family transposase
group_5008	Alpha-ketoglutaric semialdehyde dehydrogenase / NADP-dependent aldehyde dehydrogenase
group_5198	Aminoglycoside phosphotransferase
group_5212	NADH:flavin oxidoreductase
group_5385	Aminoglycoside phosphotransferase
group_5540	DUF3307 domain-containing protein
group_5542	Phosphohydrolase
group_5643	ATP/GTP-binding protein
group_5772	Tetratricopeptide repeat protein
group_5776	Radical SAM protein
group_5789	Darcynin
group_5793	Amine oxidase, flavin-containing
group_5796	Transcriptional regulator, IclR family
group_5798	Nuclear transport factor 2 family protein
group_5803	Nuclear transport factor 2 family protein
group_5818	Acyltransferase 3
group_5819	Transcriptional regulator PadR-like family protein
group_5821	ABC transporter permease
group_5836	Acyltransferase
<i>hmgA</i> <sup>a</sup>	Homogentisate 1,2-dioxygenase

<i>hsdA</i> <sup>*</sup>	3-alpha-hydroxysteroid dehydrogenase/carbonyl reductase / SDR family oxidoreductase
<i>htpG</i>	Chaperone protein HtpG / heat shock protein 90
<i>liaS</i> <sup>a</sup>	HPK7 family sensor histidine kinase LiaS
<i>mftC</i> <sup>b</sup>	Putative mycofactocin radical SAM maturase MftC / radical SAM protein
<i>ndx1</i> <sup>*</sup>	NUDIX hydrolase
<i>nuoA</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit A
<i>nuoH</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit H
<i>nuoJ</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit J
<i>nuoK</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit K
<i>nuoL</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit L
<i>nuoM</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit M
<i>nuoN</i> <sup>a</sup>	NADH-quinone oxidoreductase subunit N
<i>proP</i> <sup>a</sup>	Proline/betaine transporter
<i>ptsG</i>	PTS system glucose-specific EIICB component
<i>rhmR</i> <sup>a</sup>	HTH-type transcriptional regulator KipR / MarR family transcriptional regulator
<i>scoA</i> <sup>*</sup>	3-oxoacid CoA-transferase, A subunit
<i>scoB</i> <sup>a</sup>	3-oxoacid CoA-transferase, B subunit
<i>ssuE</i> <sup>b</sup>	FMN reductase (NADPH)
<i>tauB</i> <sup>a</sup>	Aliphatic sulfonates import ATP-binding protein SsuB / ABC transporter ATP-binding protein
<i>tcrA</i> <sup>a</sup>	Transcriptional regulatory protein CutR / DNA-binding response regulator
<i>xecD</i> <sup>a</sup>	SDR family oxidoreductase
<i>yihX</i> <sup>*</sup>	Alpha-D-glucose 1-phosphate phosphatase YihX / HAD family phosphatase
<i>yknY</i> <sup>a</sup>	Uncharacterised ABC transporter ATP-binding protein YknY

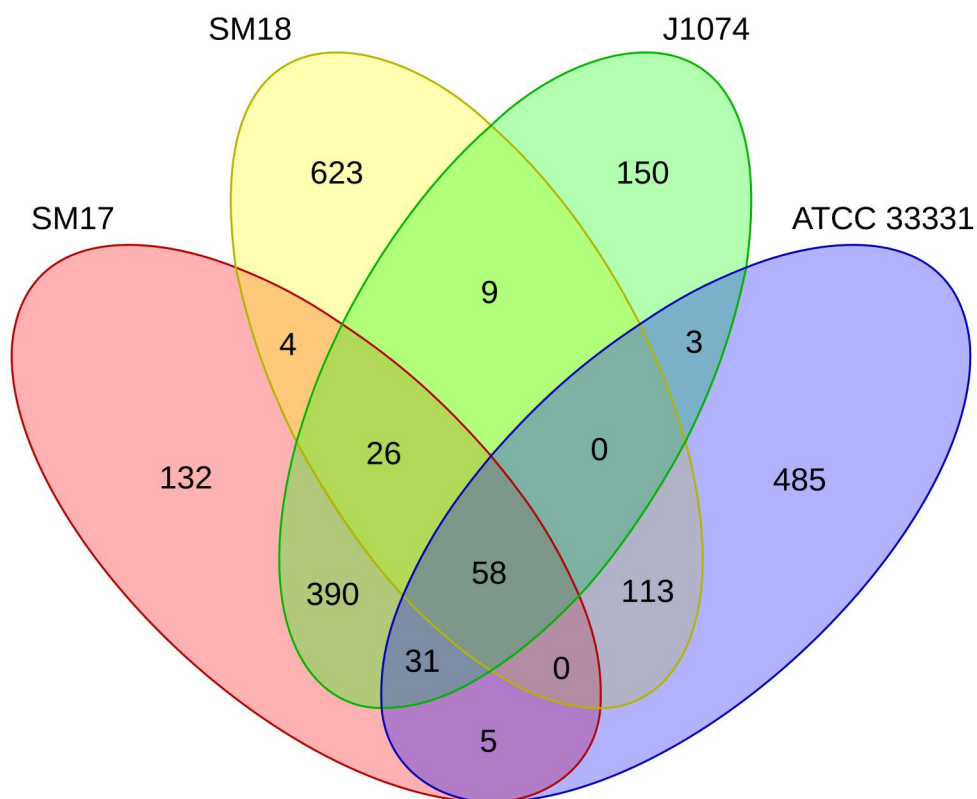
**Table S5:** List of orthologous genes and their respective annotations (excluding hypothetical proteins), which are commonly present in the sponge-derived isolates SM17 and SM18, while absent in their terrestrial counterparts J1074 and ATCC 33331. When the gene name was not determined, a generic unique name was given (group\_XXXX) by the Roary program. '<sup>\*</sup>' - genes without multiple copies or paralogs in the terrestrial isolates' genomes, considering only the ones with a defined gene name; '<sup>a</sup>' - genes for which both sponge-derived isolates presented a higher copy number in comparison to their terrestrial counterparts; '<sup>b</sup>' - genes for which the SM17 isolate had a higher copy number in comparison to its terrestrial counterpart *S. albus* J1074; '<sup>c</sup>' - genes for which the SM18 isolate had a higher copy number in comparison to its terrestrial counterpart *S. pratensis* ATCC 33331.

	<i>Streptomyces</i> sp. SM17	<i>Streptomyces</i> sp. SM18	<i>Salinispora</i> <i>arenicola</i> CNS-205	<i>Salinispora</i> <i>tropica</i> CNB-440	<i>Kocuria</i> <i>flava</i> S43
<i>Streptomyces</i> sp. SM17	-	70%	69%	69%	70%
<i>Streptomyces</i> sp. SM18	70%	-	68%	69%	68%
<i>Salinispora</i> <i>arenicola</i> CNS-205	69%	68%	-	86%	73%
<i>Salinispora</i> <i>tropica</i> CNB- 440	69%	69%	86%	-	73%
<i>Kocuria flava</i> S43	70%	68%	73%	73%	-

2837 **Table S6:** Nucleotide sequence identity comparison between the  
2838 partial *nuo*-operon present in the marine isolates *Streptomyces* sp.  
2839 SM17, *Streptomyces* sp. SM18, *Salinispora arenicola* CNS-205,  
2840 *Salinispora tropica* CNB-440, and *Kocuria flava* S43, obtained using  
2841 discontinuous MegaBLAST alignments.



**Figure S1:** Whole genomes nucleotide alignments performed using the MUMmer 3 program. A) Alignment between the *S. albus* J1074 chromosome (x axis) and the SM17 chromosome (y axis) sequences. B) Alignment between the *S. pratensis* ATCC 33331 chromosome (x axis) and the SM18 chromosome (y axis) sequences.



**Figure S2:** Venn diagram representing the presence/absence of orthologous genes in the genome regions predicted to contain smBGCs in the SM17, SM18, *S. albus* J1074, and the *S. pratensis* ATCC 33331 genomes.

# Chapter 3

Genome mining coupled with OSMAC-based cultivation reveal differential production of surugamide A by the marine sponge isolate *Streptomyces* sp. SM17 when compared to its terrestrial relative *S. albidoflavus* J1074

Eduardo L. Almeida <sup>1</sup>, Navdeep Kaur <sup>2</sup>, Laurence K. Jennings <sup>2</sup>, Andrés Felipe Carrillo Rincón <sup>1</sup>, Stephen A. Jackson <sup>1,3</sup>, Olivier P. Thomas <sup>2</sup> and Alan D.W. Dobson <sup>1,3</sup>

<sup>1</sup>School of Microbiology, University College Cork, T12YN60 Cork, Ireland

<sup>2</sup>Marine Biodiscovery, School of Chemistry and Ryan Institute, National University of Ireland Galway (NUI Galway), University Road, H91 TK33 Galway

The work presented in this chapter has been published in the journal

*Microorganisms*, 26 September 2019,

<https://doi.org/10.3390/microorganisms7100394>.

**Citation:** Almeida E. L., Kaur N., Jennings L. K., Carrillo Rincón A. F., Jackson S. A., Thomas O. P., Dobson A. D. W. (2019). Genome Mining Coupled with OSMAC-Based Cultivation Reveal Differential Production of Surugamide A by the Marine Sponge Isolate *Streptomyces* sp. SM17 When Compared to Its Terrestrial Relative *S. albidoflavus* J1074. *Microorganisms* 7, 394. doi:10.3390/microorganisms7100394.



2851    **1.    Abstract**

2852            Much recent interest has arisen in investigating *Streptomyces* isolates  
2853    derived from the marine environment in the search for new bioactive compounds,  
2854    particularly those found in association with marine invertebrates, such as sponges.  
2855    Among these new compounds recently identified from marine *Streptomyces*  
2856    isolates are the octapeptidic surugamides, which have been shown to possess  
2857    anticancer and antifungal activities. In this study, based on genome mining  
2858    followed by an OSMAC-based approach, we identified the previously unreported  
2859    capability of a marine sponge-derived isolate, namely *Streptomyces* sp. SM17, to  
2860    produce surugamide A. Phylogenomics analyses provided novel insights with  
2861    respect to the distribution and conservation of the surugamides biosynthetic gene  
2862    cluster (*sur* BGC) at a genetic level. We observed differential production of  
2863    surugamide A when comparing the closely related marine and terrestrial isolates,  
2864    namely *Streptomyces* sp. SM17 and *Streptomyces albidoflavus* J1074. SM17  
2865    produced higher levels of surugamide A than *S. albidoflavus* J1074 under all  
2866    conditions tested, and in particular producing >13-fold higher levels when grown  
2867    in YD and 3-fold higher levels in SYP-NaCl medium. In addition, surugamide A  
2868    production was repressed in TSB and YD medium, suggesting that carbon  
2869    catabolite repression (CCR) may influence the production of surugamides in these  
2870    strains.

## 2. Introduction

Members of the *Streptomyces* genus are widely known to be prolific producers of natural products. Many of these compounds have found widespread use in the pharmaceutical industry as antibiotics, immunosuppressant, antifungal, anticancer and anti-parasitic drugs (Hwang et al., 2014). However, there continues to be an urgent need to discover new bioactive compounds, and especially antibiotics; primarily due to the emergence of antibiotic resistance in clinically important bacterial pathogens (Thabit et al., 2015; Tommasi et al., 2015). In particular, the increase in multi-resistant ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) has focused research efforts to develop new antibiotics to treat these priority antibiotic-resistant bacteria (Demers et al., 2018).

Up until relatively recently, marine ecosystems had largely been neglected as a potential source for the discovery of novel bioactive compounds, in comparison to terrestrial environments; primarily due to issues of accessibility (Indraningrat et al., 2016). Marine sponges are known to host a variety of different bacteria and fungi which produce a diverse range of natural products, including compounds with antiviral, antifungal, antiprotozoal, antibacterial, and anticancer activities (Calcabrini et al., 2017; Indraningrat et al., 2016). Marine sponge-associated *Streptomyces* spp. are a particularly important source of bioactive compounds, with examples including *Streptomyces* sp. HB202, isolated from the sponge *Halichondria panicea* which produces mayamycin, a compound with activity against *Staphylococcus aureus* (Schneemann et al., 2010); and streptophenazines G and K, with activity against *Bacillus subtilis* (Kunz et al., 2014); together with *Streptomyces* sp. MAPS15, which was isolated from *Spongia officinalis* which produces 2-pyrrolidine, with activity against *Klebsiella*

2898 *pneumoniae* (Sathiyarayanan et al., 2014). Additionally, our group has reported  
2899 the production of antimycins from *Streptomyces* sp. SM8 isolated from the sponge  
2900 *Haliclona simulans*, with antifungal and antibacterial activities (Almeida et al.,  
2901 2018; Viegelmann et al., 2014). In further work we genetically characterised 13  
2902 *Streptomyces* spp. that were isolated from both shallow and deep-sea sponges,  
2903 which displayed antimicrobial activities against a number of clinically relevant  
2904 bacterial and yeast species (Jackson et al., 2018; Kennedy et al., 2009). Amongst  
2905 these strains, the *Streptomyces* sp. SM17 demonstrated an ability to inhibit the  
2906 growth of *E. coli* NCIMB 12210, methicillin resistant *S. aureus* (MRSA) and  
2907 *Candida* spp., when employing deferred antagonism assays (Jackson et al., 2018;  
2908 Kennedy et al., 2009).

2909       Among other clinically relevant natural products derived from marine  
2910 *Streptomyces* isolates are the recently identified surugamides family of molecules.  
2911 The cyclic octapeptide surugamide A and its derivatives were originally identified  
2912 in the marine-derived *Streptomyces* sp. JAMM992 (Takada et al., 2013), and have  
2913 been shown to belong to a particularly interesting family of compounds due not  
2914 only to their relevant bioactivity, but also due to their unusual metabolic pathway  
2915 involving d-amino acids (Matsuda et al., 2019b; Takada et al., 2013; Xu et al., 2017).  
2916 Since their discovery, concerted efforts have been employed in order to chemically  
2917 characterise these compounds and determine the genetic mechanisms involved in  
2918 their production (Kuranaga et al., 2018; Matsuda et al., 2019b; Ninomiya et al., 2016;  
2919 Takada et al., 2013; Thankachan et al., 2019; Zhou et al., 2019). The surugamides  
2920 and their derivatives have been shown to possess a number of bioactivities, with  
2921 the surugamides A-E and the surugamides G-J being shown to possess anticancer  
2922 activity by inhibiting bovine cathepsin B, a cysteine protease reported to be  
2923 involved in the invasion of metastatic tumour cells (Takada et al., 2013; Xu et al.,  
2924 2017); while another derivative, namely acyl-surugamide A, has been shown to

possess anti-fungal activity (Xu et al., 2017). It has been determined that the non-ribosomal peptide synthase-encoding *surABCD* genes are the main biosynthetic genes involved in the biosynthesis of surugamides and their derivatives (Ninomiya et al., 2016), with these genes being involved in the production of at least 20 different compounds (Xu et al., 2017). Surugamides A-E have been reported to be produced by the *surA* and *surD* genes, while the linear decapeptide surugamide F has been shown to be produced by the *surB* and *surC* genes, involving a unique pattern of intercalation of the biosynthetic genes (Ninomiya et al., 2016). Further metabolic pathways studies have reported that the expression of the *surABCD* gene cluster is strongly regulated by the *surR* transcriptional repressor (Xu et al., 2017), while the cyclisation of the cyclic surugamides has been shown to involve a penicillin binding protein (PBP)-like thioesterase encoded by the *surE* gene (Matsuda et al., 2019a; Thankachan et al., 2019; Zhou et al., 2019).

Although apparently widespread in marine-derived *Streptomyces* isolates (Ninomiya et al., 2016; Zhou et al., 2019), the production of surugamides has also been reported in the *S. albidoflavus* strain J1074 (Koshla et al., 2019; Xu et al., 2017), a derivative of the soil isolate *S. albus* G (Chater and Wilde, 1976, 1980). The *S. albidoflavus* strain J1074 is a well characterised *Streptomyces* isolates which is frequently used as a model for the genus, and has commonly been successfully employed in the heterologous expression of biosynthetic gene clusters (BGCs) (Bilyk et al., 2016; Huang et al., 2019; Jiang et al., 2018; Myronovskyi et al., 2018; Zaburannyi et al., 2014). This strain was originally classified as an *S. albus* isolate, however, due to more recent taxonomy studies, it has been re-classified as a *S. albidoflavus* species isolate (Labeda et al., 2014, 2017). Interestingly, surugamides and their derivatives have been shown to only be produced by *S. albidoflavus* J1074 under specific conditions, such as when employing chemical stress elicitors

(Xu et al., 2017), and more recently when cultivating the strain in a soytone-based liquid based medium SG2(Koshla et al., 2019).

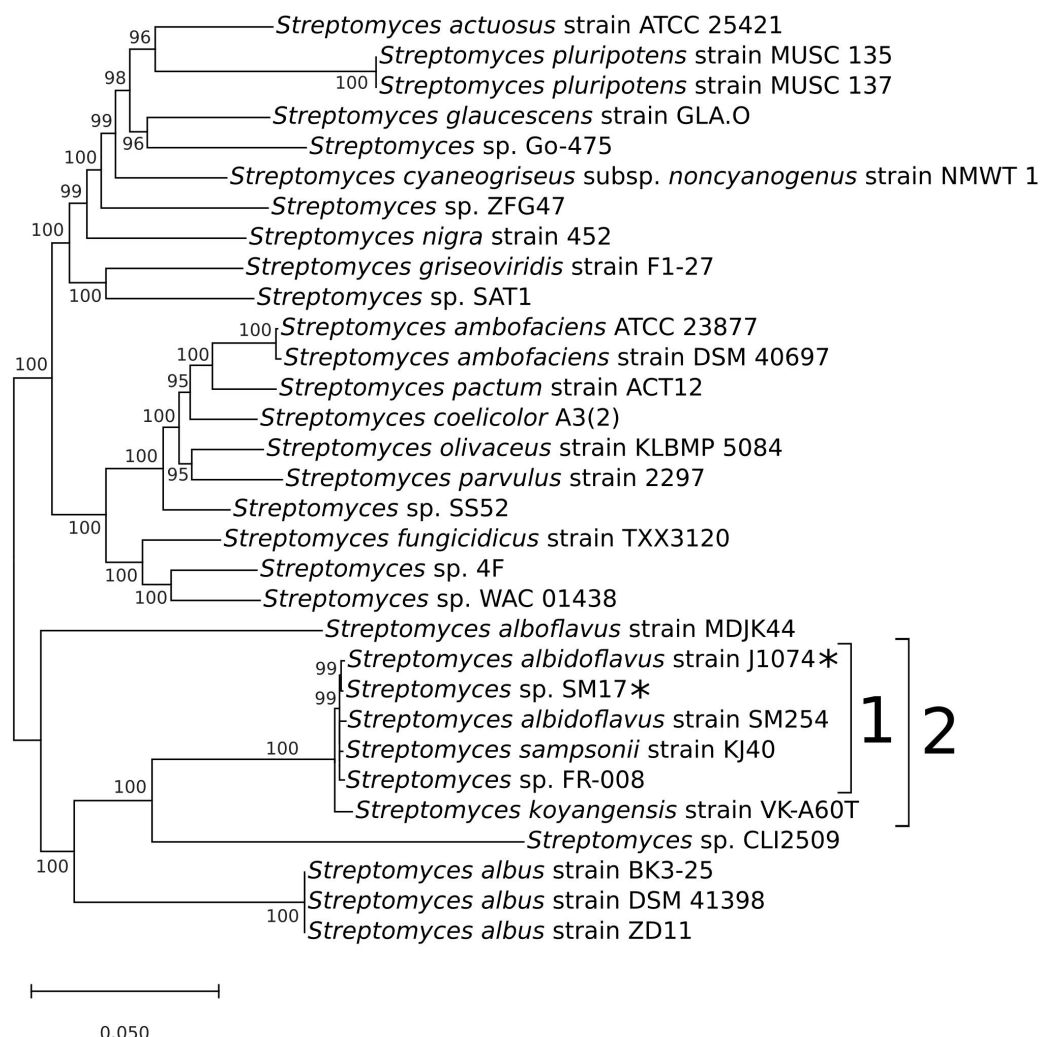
In a previous study (Almeida et al., 2019), we reported that the *S. albidoflavus* J1074 and *Streptomyces* sp. SM17 possessed morphological and genetic similarities. Differences were observed, however, when both strains were exposed to high salt concentrations using culture media, such as TSB or ISP2, in which the marine sponge-derived strain SM17 grew and differentiated more rapidly in comparison with the soil strain *S. albidoflavus* J1074, which appeared to have trouble growing and differentiating when salts were present in the growth medium (Almeida et al., 2019). Genome mining based on the prediction of secondary metabolites BGCs also showed many similarities between the two strains (Almeida et al., 2019). Among these predicted BGCs, both the *S. albidoflavus* J1074 and *Streptomyces* sp. SM17 isolates appeared to possess the *sur* BGC, encoding for the production of surugamides A/D; which had previously been identified in other marine-derived *Streptomyces* isolates (Ninomiya et al., 2016; Zhou et al., 2019), and which as has been previously mentioned is produced by the terrestrial isolate *S. albidoflavus* J1074 under specific culture conditions (Koshla et al., 2019; Xu et al., 2017). This prompted us to further investigate other isolates that shared genetic similarities to *S. albidoflavus* J1074, with respect to potential similarities in their genomes and between the metabolic profiles of the marine *Streptomyces* sp. SM17 and terrestrial *S. albidoflavus* isolates; particularly with respect to production of the well characterised *sur* BGC natural product surugamide A. To this end we employed an "One Strain Many Compounds" (OSMAC)-based approach, which has been shown to be a useful strategy in eliciting production of natural products from silent gene clusters by employing different culture conditions (Pan et al., 2019; Romano et al., 2018); to monitor production of surugamide A in both *S. albidoflavus* J1074 and *Streptomyces* sp. SM17.

### 2978 3. Results and discussion

#### 2979 3.1. Multi-locus sequence analysis and taxonomy assignment of the *Streptomyces* 2980 sp. SM17 isolate

2981 In order to taxonomically characterise the *Streptomyces* sp. SM17 isolate  
2982 based on genetic evidence, multi-locus sequence analysis (MLSA) (Glaeser and  
2983 Kämpfer, 2015) employing the 16S rRNA sequence, in addition to five  
2984 housekeeping genes, namely *atpD* (ATP synthase subunit beta), *gyrB* (DNA gyrase  
2985 subunit B), *recA* (recombinase RecA), *rpoB* (DNA-directed RNA polymerase  
2986 subunit beta), and *trpB* (tryptophan synthase beta chain) was performed, in a  
2987 similar manner to a previous report (Almeida et al., 2019). A similarity search was  
2988 performed in the GenBank database (Benson et al., 2018), using the NCBI BLASTN  
2989 tool (Camacho et al., 2009; Johnson et al., 2008), based on the 16S rRNA nucleotide  
2990 sequence of the SM17 isolate. The top 30 most similar *Streptomyces* species for  
2991 which complete genome sequences were available in GenBank were selected for  
2992 further phylogenetic analysis.

2993 The concatenated nucleotide sequences (Gadagkar et al., 2005; Glaeser and  
2994 Kämpfer, 2015) of the 16S rRNA and the aforementioned five housekeeping genes,  
2995 were first aligned using the MAFFT program (Kato and Standley, 2013), and the  
2996 phylogeny analysis was performed using the MrBayes program (Ronquist et al.,  
2997 2012). The General Time Reversible (GTR) model of nucleotide substitution with  
2998 gamma-distributed rates across sites with a proportion of invariable sites was  
2999 applied (Waddell and Steel, 1997), with 1 million generations sampled every 100  
3000 generations. The final phylogenetic tree was then processed using MEGA X  
3001 (Kumar et al., 2018), with a posterior probability cut off of 95% (Figure 1).



**Figure 1:** Phylogenetic tree of the concatenated sequences of the 16S rRNA and the housekeeping genes *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB*, from the *Streptomyces* sp. SM17 together with 30 *Streptomyces* isolates for which complete genome sequences were available in the GenBank database. Analysis was performed using MrBayes, with a posterior probability cut off of 95%. 1) *albidoflavus* phylogroup. 2) clade including the neighbour isolate *Streptomyces koyangensis* strain VK-A60T. The strains SM17 and J1074 are indicated with asterisks.

3011           The resulting phylogenetic tree clearly indicates the presence of a clade that  
 3012 includes the isolates *Streptomyces albidoflavus* strain J1074; *Streptomyces* sp.  
 3013 SM17; *Streptomyces albidoflavus* strain SM254; *Streptomyces sampsonii* strain  
 3014 KJ40; *Streptomyces* sp. FR-008; and *Streptomyces koyangensis* strain VK-A60T  
 3015 (clade 2 in Figure 1). In addition, this larger clade contains a sub-clade (clade 1 in  
 3016 Figure 1) that includes *Streptomyces* isolates similar to the type-strain  
 3017 *Streptomyces albidoflavus* J1074. The J1074 strain is a well-studied *Streptomyces*  
 3018 isolate widely used as a model for the genus and for various biotechnological  
 3019 applications, including the heterologous expression of secondary metabolites  
 3020 biosynthetic gene clusters (BGCs) (Bilyk et al., 2016; Huang et al., 2019; Jiang et al.,  
 3021 2018; Myronovskyi et al., 2018; Zaburannyi et al., 2014). This isolate was originally  
 3022 classified as “*Streptomyces albus* J1074”, but due to recent taxonomy data, it has  
 3023 been re-classified as *Streptomyces albidoflavus* J1074 (Labeda et al., 2014, 2017).  
 3024 Hence, in this study, this strain will be referred to as *Streptomyces albidoflavus*  
 3025 J1074, and this clade will from now on be referred to as the *albidoflavus*  
 3026 phylogroup (Figure 1).

3027           Interestingly, members of the *albidoflavus* phylogroup were all isolated  
 3028 from quite different environments. The *Streptomyces albidoflavus* strain J1074  
 3029 stems from the soil isolate *Streptomyces albus* G (Chater and Wilde, 1976, 1980).  
 3030 The *Streptomyces sampsonii* strain KJ40 was isolated from rhizosphere soil in a  
 3031 poplar plantation (Li et al., 2018). The *Streptomyces* sp. strain FR-008 is a random  
 3032 protoplast fusion derivative of two *Streptomyces hygroscopicus* isolates (Liu et al.,  
 3033 2016). On the other hand, two of these strains were isolated from aquatic saline  
 3034 environments, with *Streptomyces* sp. SM17 being isolated from the marine sponge  
 3035 *Haliclona simulans* (Kennedy et al., 2009); while the *Streptomyces albidoflavus*  
 3036 strain SM254 strain was isolated from copper-rich subsurface fluids within an iron  
 3037 mine, following growth on artificial sea water (ASW) (Badalamenti et al., 2016). The



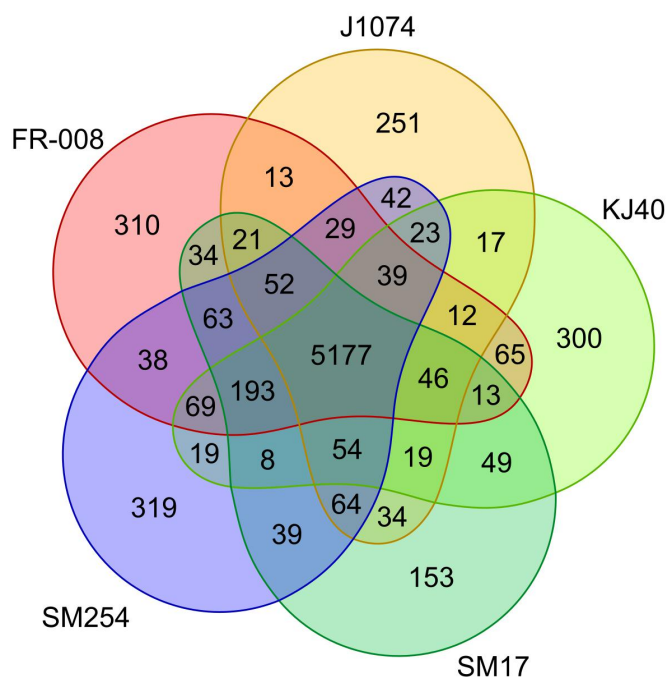
fact that these isolates, although derived from quite distinct environmental niches, simultaneously share significant genetic similarities is interesting, and raises questions about their potential evolutionary relatedness.

### 3.2. Analysis of groups of orthologous genes in the *albidoflavus* phylogroup

In an attempt to provide further genetic evidence with respect to the similarities shared among the members of the *albidoflavus* phylogroup (Figure 1), a pan-genome analysis was performed to determine the number of core genes, accessory genes, and unique genes present in this group of isolates. The Roary program was employed for this objective (Page et al., 2015), which allowed the identification of groups of orthologous and paralogous genes (which from now on will be referred to simply as “genes”) present in the set of *albidoflavus* genomes, with a protein identity cut off of 95%, which is the identity value recommended by the Roary program manual when analysing organisms belonging to the same species.

A total of 7,565 genes were identified in the *albidoflavus* pan-genome, and among these a total of 5,177 were determined to be shared among all the *albidoflavus* isolates (i.e. the core genome) (Figure 2). This represents a remarkably high proportion of genes which appear to be highly conserved between all the isolates, representing approximately 68.4% of the pan-genome. Additionally, when considering the genomes individually (Table S1), the core genome accounts for approximately 84.5% of the FR-008 genome; 88.5% of J1074; 85.5% of KJ40; 86.7% of SM17; and 83.7% of the SM254 genome. On the other hand, the accessory genome (i.e. genes present in at least two isolates) was determined to consist of 1,055 genes (or ~13.9% of the pan-genome); while the unique genome (i.e. genes present in only one isolate) was determined to comprise of 1,333 genes (or ~17.6% of the pan-genome). This strikingly high conservation of genes present in their genomes,

3064 together with the previous multi-locus phylogeny analysis, are very strong  
 3065 indicators that these microorganisms may belong to the same species.



3066 **Figure 2:** Venn diagram representing the presence/absence of groups  
 3067 of orthologous genes in the organisms belonging to the *albidoflavus*  
 3068 phylogroup.

3069 An additional pan-genome analysis similar to the aforementioned analysis  
 3070 was also performed including the *Streptomyces koyangensis* strain VK-A60T in the  
 3071 dataset (Figure S1), which was an isolate shown to be a closely related neighbour to  
 3072 the *albidoflavus* phylogroup (Figure 1, clade 2). When compared to the previous  
 3073 analysis, the pan-genome analysis including the VK-A60T isolate showed  
 3074 significant changes in the values representing the core genome, which changed  
 3075 from 5,177 genes (Figure 2) to 3,912 genes (Figure S1), with an additional 1,273  
 3076 genes also shared among all of the *albidoflavus* isolates (Figure S1). The results also  
 3077 showed a much larger number of genes uniquely present in the VK-A60T genome

3078 than in the other genomes, with 2,059 unique genes identified from a total of 6,245  
3079 CDSs present in the VK-A60T genome in total, or approximately a third of its total  
3080 number of genes (Figure S1). This proportion of unique genes present in the VK-  
3081 A60T genome is considerably higher than the proportions of unique genes  
3082 observed in the other *albidoflavus* phylotype genomes (Figure 2), which accounted  
3083 for approximately only 2.5% of the total number of genes in SM17; 4.2% in J1074;  
3084 4.9% in KJ40; 5% in FR-008; and 5.1% in SM254. Taken together, these results  
3085 further demonstrate the similarities between the isolates belonging to the  
3086 *albidoflavus* phylogroup, while the VK-A60T isolate is clearly more distantly  
3087 related.

3088         Thus, from previous studies (Labeda et al., 2014, 2017), and in light of the  
3089 phylogeny analysis and further genomic evidence presented in this study, it is  
3090 likely that all the isolates belonging to the *albidoflavus* phylogroup are in fact  
3091 members of the same species. It is reasonable to infer that, for example, the isolates  
3092 in the *albidoflavus* phylogroup that possess no species assignment thus far (i.e.  
3093 strains SM17 and FR-008) are indeed members of the *albidoflavus* species. Also, it  
3094 is possible that the *Streptomyces sampsonii* KJ40 has been misassigned, and  
3095 possibly requires re-classification as an *albidoflavus* isolate.

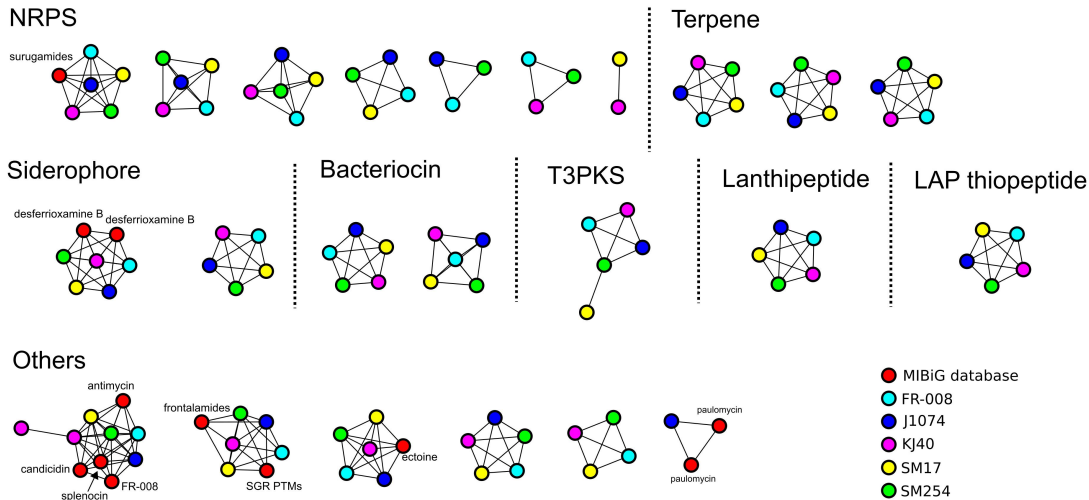
3096         Misassignment and re-classification of *Streptomyces* species is a common  
3097 issue, and an increase in the quantity and the quality of available data from these  
3098 organisms (e.g. better quality genomes available in the databases) will provide  
3099 better support for taxonomy claims, or correction of these when new information  
3100 becomes available (Labeda et al., 2017; Li et al., 2019; Rong et al., 2013; Ward and  
3101 Allenby, 2018).

### 3.3. Prediction of secondary metabolites biosynthetic gene clusters in the *albidoflavus* phylogroup

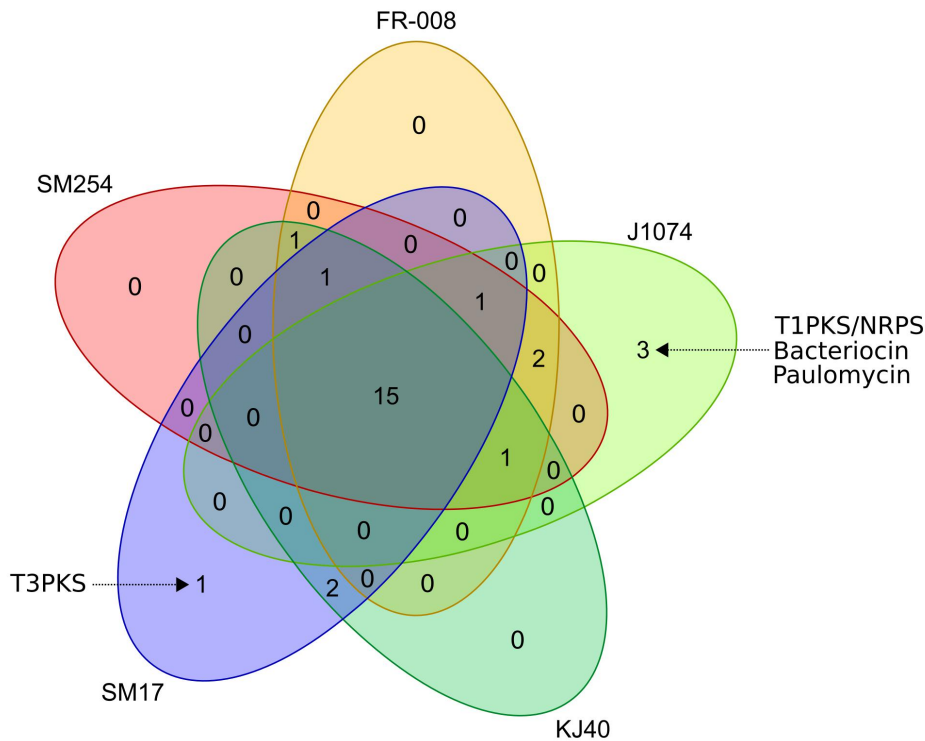
Isolates belonging to the *albidoflavus* phylogroup have been reported to produce bioactive compounds of pharmacological relevance, such as antibiotics. As mentioned previously, the *Streptomyces albidoflavus* strain J1074 is the best described member of the *albidoflavus* phylogroup to date. As such, several of secondary metabolites produced by this isolate have been identified, including acyl-surugamides and surugamides with antifungal and anticancer activities, respectively (Xu et al., 2017); together with paulomycin derivatives with antibacterial activity (Hoz et al., 2017). The *Streptomyces* sp. FR-008 isolate has been shown to produce the antimicrobial compound FR-008/candididin (Chen et al., 2003; Zhao et al., 2015); while the *Streptomyces sampsonii* KJ40 isolate has been shown to produce a chitinase that possesses anti-fungal activity against plant pathogens (Li et al., 2018). On the other hand, although no bioactive compound have been characterised from *Streptomyces albidoflavus* SM254, this isolate has been shown to possess anti-fungal activity, specifically against the fungal bat pathogen *Pseudogymnoascus destructans*, which is responsible for the White-nose Syndrome (Badalamenti et al., 2016; Hamm et al., 2017). The *Streptomyces* sp. SM17 isolate has also previously been shown to possess antibacterial and antifungal activities against clinically relevant pathogens, including methicillin-resistant *Staphylococcus aureus* (MRSA) (Kennedy et al., 2009). However, no natural products derived from this strain have been identified and isolated until now.

In order to further *in silico* assess the potential of these *albidoflavus* phylogroup isolates to produce secondary metabolites, and also to determine how potentially similar or diverse they are within this phylogroup, prediction of secondary metabolites biosynthetic gene clusters (BGCs) was performed using the

3129 antiSMASH (version 5) program (Blin et al., 2019). The antiSMASH prediction was  
3130 processed using the BiG-SCAPE program (Navarro-Muñoz et al., 2018), in order to  
3131 cluster the BGCs into gene cluster families (GCFs), based on sequence and Pfam  
3132 (El-Gebali et al., 2019) protein families similarity, and also by comparing them to  
3133 the BGCs available from the Minimum Information about a Biosynthetic Gene  
3134 cluster (MIBiG) repository (Medema et al., 2015) (Figure 3). When compared to  
3135 known BGCs from the MIBiG database, a significant number of BGCs predicted to  
3136 be present in the *albidoflavus* phylogroup genomes could potentially encode for  
3137 the production of novel compounds, including those belonging to the non-  
3138 ribosomal peptide synthetase (NRPS) and bacteriocin families of compounds  
3139 (Figure 3). The presence/absence of homologous BGCs in the *albidoflavus* isolates'  
3140 genomes was determined using BiG-SCAPE and is represented in Figure 4.  
3141 Interestingly, the vast majority of the BGCs predicted in the *albidoflavus*  
3142 phylogroup are shared among all of its members (15 BGCs); while another large  
3143 portion (8 BGCs) are present in at least two isolates (Figure 4). Among the five  
3144 members of the *albidoflavus* phylogroup, only the J1074 strain and the SM17 strain  
3145 appeared to possess unique BGCs when compared to the other strains. Three  
3146 unique BGCs were predicted to be present in the J1074 genome: a predicted type I  
3147 polyketide synthase (T1PKS)/NRPS without significant similarity to the BGCs from  
3148 the MIBiG database; a predicted bacteriocin which also did not show any  
3149 significant similarity to the BGCs from the MIBiG database; and a BGC predicted to  
3150 encode for the production of the antibiotic paulomycin, with similarity to the  
3151 paulomycin-encoding BGCs from *Streptomyces paulus* and *Streptomyces* sp. YN86  
3152 (Li et al., 2015), which has also been experimentally shown to be produced by the  
3153 J1074 strain (Hoz et al., 2017). One BGC predicted to encode a type III polyketide  
3154 synthase (T3PKS) – with no significant similarity to the BGCs from the MIBiG  
3155 database – was also identified as being unique to the SM17 genome.



**Figure 3:** BGCs similarity clustering using BiG-SCAPE. Singletons, i.e. BGCs without significant similarity with the BGCs from the MIBiG database or with the BGCs predicted in other genomes, are not represented.



**Figure 4:** Venn diagram representing BGCs presence/absence in the genomes of the members of the *albidoflavus* phylogroup, determined using antiSMASH and BiG-SCAPE.

Importantly, BGCs with similarity to the surugamide A/D BGC from “*Streptomyces albus* J1074” (now classified as *S. albidoflavus*) from the MIBiG database (Xu et al., 2017) were identified in all the other genomes of the members of the *albidoflavus* phylogroup. This raises the possibility that this BGC may be commonly present in *albidoflavus* species isolates. However, as only a few complete genomes of isolates belonging to this phylogroup are currently available, further data will be required to support this hypothesis. Nevertheless, these results further highlight the genetic similarities of the isolates belonging to the *albidoflavus* phylogroup, even with respect to their potential to produce secondary metabolites.

#### 3.4. Phylogeny and gene synteny analysis of *sur* BGC homologs

In parallel to the previous phylogenomics analysis performed with the *albidoflavus* phylogroup isolates, sequence similarity and phylogenetic analyses were performed, using the previously described and experimentally characterised *Streptomyces albidoflavus* LHW3101 surugamides biosynthetic gene cluster (*sur* BGC, GenBank accession number: MH070261) as a reference (Zhou et al., 2019). The aim was to assess how widespread in nature the *sur* BGC might be, and the degree of genetic variation, if any; that might be present in *sur* BGCs belonging to different microorganisms.

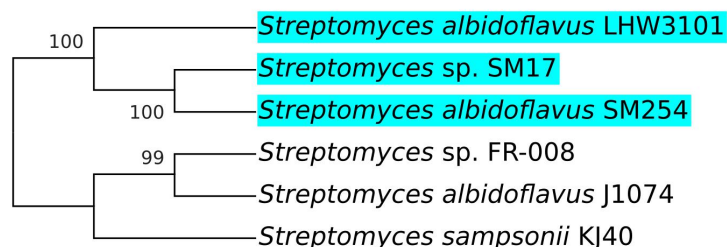
Nucleotides sequence similarity to the *sur* BGC was performed in the GenBank database (Benson et al., 2018), using the NCBI BLASTN tool (Camacho et al., 2009; Johnson et al., 2008). It is important to note that, since the quality of the data is crucial for sequence similarity, homology and phylogeny inquiries; only complete genome sequences were employed in this analysis. For this reason, for example, the marine *Streptomyces* isolate in which surugamides and derivatives were originally identified, namely *Streptomyces* sp. JAMM992 (Takada et al., 2013),

3188 was not included, since its complete genome is not available in the GenBank  
3189 database.

3190 The sequence similarity analysis identified 5 microorganisms that possessed  
3191 homologs to the *sur* BGC and had their complete genome sequences available in  
3192 the GenBank database: *Streptomyces* sp. SM17; *Streptomyces albidoflavus* SM254;  
3193 *Streptomyces* sp. FR-008; *Streptomyces albidoflavus* J1074; and *Streptomyces*  
3194 *sampsonii* KJ40. Notably, these results overlapped with the isolates belonging to  
3195 the previously discussed *albidoflavus* phylogroup (Figure 1), further highlighting  
3196 the possibility that the *sur* BGC may be commonly present in and potentially  
3197 exclusive to the *albidoflavus* species.

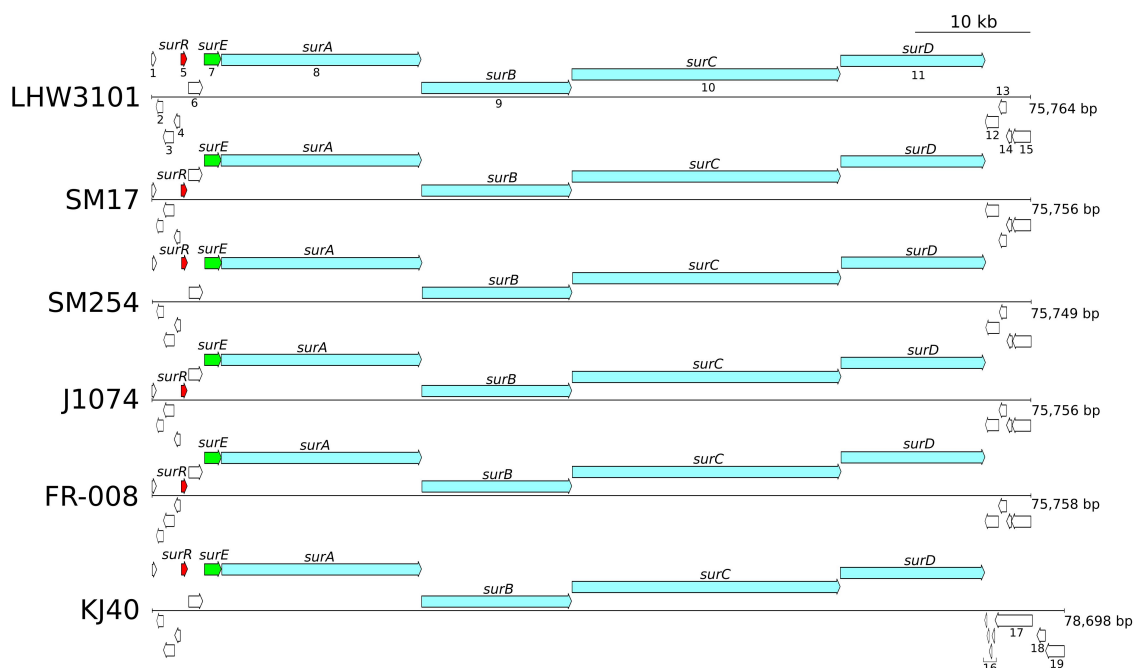
3198 Phylogenetic analysis was performed in the genomic regions determined to  
3199 be homologs to the *Streptomyces albidoflavus* LHW3101 *sur* BGC, using the  
3200 MrBayes program (Ronquist et al., 2012) (Figure 5). Although a larger number of  
3201 sequences should ideally be employed in this type of analysis, these results suggest  
3202 the possibility of a clade with aquatic saline environment-derived *sur* BGCs (Figure  
3203 5). Thus, these aquatic saline environment-derived *sur* BGCs are likely to share  
3204 more genetic similarities amongst each other, rather than with those derived from  
3205 terrestrial environments. Since this analysis took into consideration the whole  
3206 genome regions that contained the *sur* BGCs of each isolate, it is likely that the  
3207 similarities and differences present in these regions involve not only coding  
3208 sequences (CDSs) for biosynthetic genes and/or transcriptional regulators, but also  
3209 could include promoter regions and other intergenic sequences.





**Figure 5:** Consensus phylogenetic tree of the *sur* BGC region of the *S. albidoflavus* LHW3101 reference *sur* BGC sequence, plus 5 *Streptomyces* isolates determined to have *sur* BGC homologs, generated using MrBayes and Mega X, with a 95% posterior probability cut off. Aquatic saline environment-derived isolates are highlighted in cyan.

With this in mind, the genomic regions previously determined to share homology with the *sur* BGC from *S. albidoflavus* LHW3101 were further analysed, with respect to the genes present in the surrounding region, the organisation of the BGCs, together with the overall gene synteny (Figure 6). Translated CDSs predicted in the region were manually annotated using the NCBI BLASTP tool (Camacho et al., 2009; Johnson et al., 2008), together with GenBank (Benson et al., 2018) and the CDD (Marchler-Bauer et al., 2015) databases. These included the main biosynthetic genes, namely *surABCD*, the transcriptional regulator SurR, and the thioesterase SurE – all of which had previously been reported to have roles in the biosynthesis of surugamides and their derivatives (Matsuda et al., 2019b; Ninomiya et al., 2016; Thankachan et al., 2019; Xu et al., 2017; Zhou et al., 2019) (Figure 6).



**Figure 6:** Gene synteny of the *sur* BGC region, including the reference *sur* BGC nucleotide sequence (LHW3101) and each of the *albidoflavus* phylogroup genomes. Arrows at different positions represent genes transcribed in different reading frames.

Interestingly, this result indicated that the gene synteny of the biosynthetic genes as well as the flanking genes is highly conserved, with the exception to the 3' flanking region of the BGC from *S. sampsonii* KJ40. Notably, even the reading frames of the *surE* gene and the *surABCD* genes are conserved amongst all the genomes. As indicated by the numbers in Figure 6, the 5' region in all the genomic regions consisted of: 1) a MbtH-like protein, which have been reported to be involved in the synthesis of non-ribosomal peptides, antibiotics and siderophores, in *Streptomyces* species (Lautru et al., 2007; Quadri et al., 1998); 2) a putative ABC transporter, which is a family of proteins with varied biological functions, including conferring resistance to drugs and other toxic compounds (Glavinas et al., 2004; Polgar and Bates, 2005); 3) a BcrA family ABC transporter, which is a family commonly involved in peptide antibiotics resistance (Ohki et al., 2003; Podlessek et

3243 al., 1995); 4) a hypothetical protein; followed by 5) the transcriptional repressor  
 3244 SurR, which has been experimentally demonstrated to repress the production of  
 3245 surugamides (Xu et al., 2017); 6) a hypothetical membrane protein; 7) the  
 3246 thioesterase SurE, which is homologous to the penicillin binding protein, reported  
 3247 to be responsible for the cyclisation of surugamides molecules (Matsuda et al.,  
 3248 2019a); and finally 8-11) the main surugamides biosynthetic genes *surABCD*, all of  
 3249 which encode non-ribosomal peptide synthetase (NRPS) proteins (Ninomiya et al.,  
 3250 2016). The 3' flanking region consisted of: 12) a predicted multi-drug resistance  
 3251 (MDR) transporter belonging to the major facilitator superfamily (MFS) of  
 3252 membrane transport proteins (Kumar et al., 2016; Yan, 2015); 13) a predicted  
 3253 TetR/AcrR transcriptional regulator, which is a family of regulators reported to be  
 3254 involved in antibiotic resistance (Cuthbertson and Nodwell, 2013); 14) a  
 3255 hypothetical protein; and 15) another predicted MDR transporter belonging to the  
 3256 MFS superfamily. In contrast, the 3' flanking region of the KJ40 strain *sur* BGC,  
 3257 consisted of: 16) a group of four hypothetical proteins, which may represent  
 3258 pseudogene versions of the first MDR transporter identified in the other isolates  
 3259 (gene number 12 in Figure 6); 17) a predicted rearrangement hotspot (RHS) repeat  
 3260 protein, which is a family of proteins reported to be involved in mediating  
 3261 intercellular competition in bacteria (Koskiniemi et al., 2013); 18) a hypothetical  
 3262 protein; and 19) a MDR transporter belonging to the MFS superfamily, which,  
 3263 interestingly, is a homolog of protein number 15, which is present in all the other  
 3264 isolates.

3265 The conserved gene synteny observed in the *sur* BGC genomic region,  
 3266 particularly those positioned upstream of the main biosynthetic *surABCD* genes,  
 3267 together with the observation that even the reading frames of the *surE* and the  
 3268 *surABCD* genes are conserved among all the genomes analysed, coupled with the  
 3269 previous phylogenetic and pan-genome analyses, suggest the following. Firstly, it

is very likely that these strains share a common ancestry and that the *sur* BGC genes had a common origin. Secondly, there is a strong evolutionary pressure ensuring the maintenance of not only gene synteny, but also of the reading frames of the main biosynthetic genes involved in the production of surugamides. The latter raises the question of which other genes in this region may be involved in the production of these compounds, or potentially conferring mechanisms of self-resistance to surugamides in the isolates, particularly since many of the genes have predicted functions that are compatible with the transport of small molecules and with multi-drug resistance. These observations are particularly interesting considering that these strains are derived from quite varied environments and geographic locations.

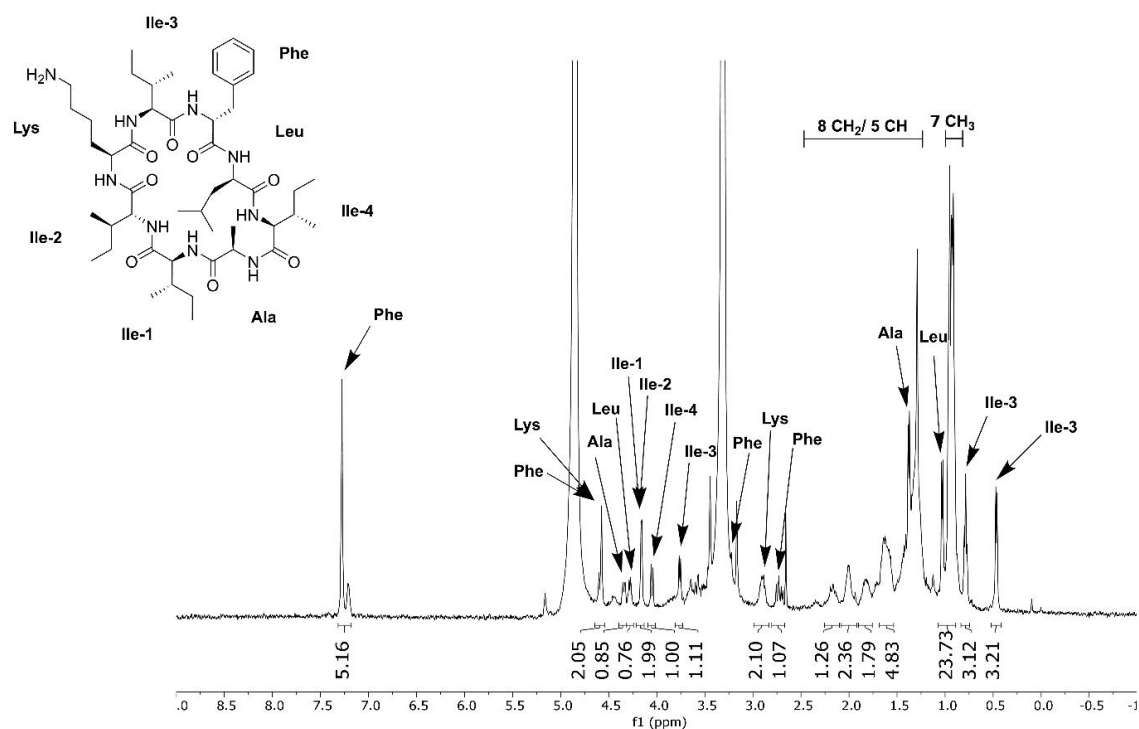
### **3.5. Growth, morphology, phenotype and metabolism assessment of *Streptomyces* sp. SM17 in complex media**

In order to assess the metabolic potential of the SM17 strain (Manteca and Yagüe, 2019), particularly with respect to the production of surugamide A, the isolate was cultivated in a number of different growth media, within an OSMAC-based approach (Pan et al., 2019; Romano et al., 2018). While the SM17 strain was able to grow in SYP-NaCl, YD, SY, P1, P2, P3 and CH-F2 liquid media; the strain was unable to grow in Oatmeal and Sporulation media. The latter indicated an inability to metabolise oat and starch when nutrients other than yeast extract are not present. Morphologically, the SM17 strain formed cell aggregates or pellets in TSB, YD and SYP-NaCl, while this differentiation was not observed in the other media. Preliminary chemical analyses of these samples, employing Liquid Chromatography–Mass Spectrometry (UPLC-DAD-HRMS), indicated that secondary metabolism in SM17 was not very active when the strain was cultivated in SY, P1, P2, P3 and CH-F2 media. In contrast, significant production of surugamides was evidenced in the extracts from TSB, SYP-NaCl and YD media,

3297 with characteristic ions at  $m/z$  934.6106 (surugamide A) and 920.5949 (surugamide  
3298 B)  $[M+Na]^+$ , which correlated with the formation of cell pellets and the production  
3299 of natural products, as previously described in other *Streptomyces* strains  
3300 (Manteca et al., 2008; Manteca and Yagüe, 2019).

3301 **3.6. Differential production of surugamide A by *Streptomyces* sp. SM17 and *S.***  
3302 ***albidoflavus* J1074**

3303 To confirm the production of surugamide A by the SM17 isolate, extracts  
3304 from the TSB, SYP-NaCl and YD media were combined and purified using High  
3305 Performance Liquid Chromatography (HPLC). The structures of the major  
3306 compounds of the extract were subsequently analysed using Nuclear Magnetic  
3307 Resonance (NMR) spectroscopy, which allowed for the identification of the  
3308 chemical structure of the surugamide A molecule as major metabolite by  
3309 comparison with reference NMR data (Figure 7) (Takada et al., 2013).



**Figure 7:** Structure of surugamide A isolated from SM17 grown in TSB, SYP-NaCl and YD medium with annotated <sup>1</sup>H NMR spectrum obtained in CD<sub>3</sub>OD at 500 MHz.

The isolates *Streptomyces* sp. SM17 and *S. albidoflavus* J1074 were subsequently cultivated in the aforementioned media in which the SM17 strain had been shown to be metabolically active, namely the TSB, SYP-NaCl, and YD media. This was performed in order to assess whether there were any significant differences in the production of surugamide A when different growth media are employed for the production of this compound, and to compare the levels of surugamide A produced by the SM17 and the J1074 isolates. The MeOH/DCM (1:1) extracts from the aforementioned cultures of SM17 and J1074 were subjected to Liquid Chromatography–Mass Spectrometry (UPLC–HRMS) to quantify the levels of surugamide A being produced under each condition (Table 1), using a surugamide A standard calibration curve (Figure S2).

Strain	Media	Percent (w/w) of Extract	Concentration of Surugamide A (mg/L) corrected in 5 mg/mL of extract
SM17	TSB	2.44%	122.01
SM17	SYP-NaCl	10.60%	530.15
SM17	YD	1.13%	56.27
J1074	TSB	0.27%	13.32
J1074	SYP-NaCl	3.55%	176.82
J1074	YD	0.09%	4.26

3324 **Table 1:** Surugamide A production by SM17 and J1074 measured  
3325 using different media.

3326 The LC-MS quantification analysis (Table 1) indicated that both strains were  
3327 capable of producing surugamide A in all the conditions tested. However, the  
3328 SM17 strain appeared to produce considerably higher yields of the compound  
3329 when compared to J1074, in all the conditions analysed. In addition, the *S.*  
3330 *albidoflavus* J1074 isolate appeared to produce quite low levels of surugamide A  
3331 when grown in TSB and YD media, accounting for less than 1% (w/w) of the  
3332 extracts from these media. Interestingly, higher yields of surugamide A were  
3333 produced in the SYP-NaCl medium in both strains, when compared with the levels  
3334 of surugamide A produced by these strains when grown in TSB and the YD media  
3335 (Table 1). In the SM17 culture in SYP-NaCl, surugamide A accounted for 10.60%  
3336 (w/w) of the extract, compared to 2.44% and 1.13% from TSB and YD, respectively;  
3337 while in J1074 it accounted for 3.55% (w/w) of the extract from the SYP-NaCl  
3338 culture, compared to 0.27% and 0.09% from TSB and YD, respectively (Table 1).  
3339 These results provide further insights into factors that are potentially involved in  
3340 regulation the biosynthesis of surugamide A, in the *albidoflavus* phylogroup and  
3341 in *Streptomyces* sp. SM17 in particular.

3342 Firstly, it appears likely that surugamide A biosynthesis may be regulated, at  
3343 least in part, by carbon catabolite repression (CCR). Carbon catabolite repression is  
3344 a well described regulatory mechanism in bacteria that controls carbon metabolism  
3345 (Brückner and Titgemeyer, 2002; Deutscher, 2008; Kremling et al., 2015; Stülke and  
3346 Hillen, 1999), and which has also been reported to regulate the biosynthesis of  
3347 secondary metabolites in a number of different bacterial species, including in  
3348 *Streptomyces* isolates (Gallo and Katz, 1972; Inoue et al., 2007; Magnus et al., 2017;  
3349 Romero-Rodríguez et al., 2016). While the TSB and the YD media contain glucose  
3350 and dextrans as carbon sources, respectively; the complex polysaccharide starch is  
3351 the carbon source in the SYP-NaCl medium. Therefore, it is reasonable to infer that  
3352 glucose and dextrin may repress the production of surugamide A in *Streptomyces*  
3353 sp. SM17 and in *Streptomyces albidoflavus* J1074, while starch does not. Further  
3354 evidence for this can be found when considering the different production media  
3355 previously employed in the production of surugamides by different *Streptomyces*  
3356 isolates. For example, in the original research that led to the discovery of  
3357 surugamides in *Streptomyces* sp. JAMM992 (Takada et al., 2013), the PC-1 medium  
3358 (1% starch, 1% polypeptone, 1% meat extract, 1% molasses, pH 7.2) was employed  
3359 for production of these compounds. Similarly to the SYP-NaCl medium employed  
3360 in our study, the PC-1 medium also contains starch as the carbon source, together  
3361 with another complex carbon source, namely molasses. Likewise, for the  
3362 production of surugamides in *S. albidoflavus* strain LHW3101 (Zhou et al., 2019),  
3363 the TSBY medium (3% tryptone soy broth, 10.3% sucrose, 0.5% yeast extract) was  
3364 employed, which utilises sucrose as its main carbon source. In contrast, when  
3365 elicitors were employed to induce the production of surugamides and their  
3366 derivatives in the J1074 strain (Xu et al., 2017), by activating the *sur* BGC which  
3367 appeared to be silent in this isolate; the R4 medium (0.5% glucose, 0.1% yeast  
3368 extract, among other non-carbon related components) was employed, which  
3369 utilises glucose as its main carbon supply, and, as shown in this study, it potentially



3370 represses the production of surugamide A. Thus, from these previous reports and  
3371 from our observations, it appears likely that CCR plays an important role in  
3372 regulating the biosynthesis of surugamides.

3373         Secondly, it is important to note the presence of salts in the form of NaCl in  
3374 the SYP-NaCl medium. As previously mentioned, genetic and phylogenetic  
3375 analyses of the *sur* BGC indicated similarities between those BGCs belonging to  
3376 aquatic saline-derived *Streptomyces* isolates (Figure 5); together with the  
3377 likelihood that these *sur* BGCs might have had a common origin. Thus, it is  
3378 plausible that this origin may have been marine, and hence the presence of salts in  
3379 the growth medium may also have an influence on the biosynthesis of surugamide  
3380 A. Different concentrations of salts in the form of NaCl in the culture medium have  
3381 also previously been shown to impact on the chemical profile of metabolites  
3382 produced in the marine-obligate bacteria *Salinispora arenicola* (Bose et al., 2015).

3383         Nevertheless, it is interesting to observe that, despite the  
3384 repression/induction of the biosynthesis of surugamide A observed when different  
3385 media were employed, the SM17 isolate is clearly produces considerably higher  
3386 amounts of surugamide A when compared to *S. albidoflavus* J1074 – reaching  
3387 yields up to >13-fold higher in the YD medium, and around 3-fold higher when  
3388 grown in the SYP-NaCl medium (Table 1).

3389    **4.    Conclusions**

3390            Marine-derived bacteria, particularly those isolated in association with  
3391 marine invertebrates, such as sponges, have been shown to be reservoirs of  
3392 bioactive molecules, including those with antibacterial, antifungal, and anticancer  
3393 activities. Among these newly identified bioactive compounds, the surugamides  
3394 and their derivatives are of particular interest due to their clinically relevant  
3395 bioactivities, i.e. anticancer and antifungal, and their original metabolic pathway.

3396            Based on genome mining, this study identified the previously unreported  
3397 capability of the marine sponge-derived isolate *Streptomyces* sp. SM17 to produce  
3398 surugamide A and also sheds new light on factors such as the carbon catabolite  
3399 repression (CCR) that may be involved in regulating production of this molecule.  
3400 Phylogenomics analysis indicated that the *sur* BGC is commonly present in  
3401 members of the proposed *albidoflavus* phylogroup, and that the *sur* BGCs present  
3402 in different isolates derived from varied environmental niches may possess a  
3403 common ancestry. Although high quality genomic data from this proposed  
3404 *albidoflavus* phylogroup is still lacking, results presented here suggest that the *sur*  
3405 BGCs derived from *Streptomyces* isolated from aquatic saline environment are  
3406 more similar to each other, when compared to those isolated from terrestrial  
3407 environments.

3408            Chemical analysis was performed in order to assess differential production  
3409 of surugamide A when comparing a marine *Streptomyces* isolate with a terrestrial  
3410 *Streptomyces* isolate, namely SM17 and J1074 strains, respectively; following an  
3411 OSMAC-based approach employing different culture media. This analysis showed  
3412 that not only the marine-derived isolate SM17 was capable of producing more  
3413 surugamide A when compared to J1074 under all the conditions tested, but also  
3414 that the biosynthesis of surugamide A is likely to be influenced by the CCR, and  
3415 potentially by the presence of salts in the growth medium. These results also

3416 highlight the importance of employing an OSMAC-based approach even when  
3417 analysing the production of known compounds, since there is a clear difference in  
3418 the yields of surugamide A obtained when employing different culture media.  
3419 Thus, it is possible to gain further insights into the production of bacterial types of  
3420 compounds by 1) discovering strains that possess a higher capability to produce  
3421 these compounds; 2) establishing optimal conditions for the biosynthesis of their  
3422 production; and 3) providing a better understanding of the genetic and regulatory  
3423 mechanisms potentially underpinning the production of these compounds.

## 3424 5. Materials and methods

### 3425 5.1. Bacterial strains and nucleotide sequences

3426 The *Streptomyces* sp. SM17 strain was isolated from the marine sponge  
3427 *Haliclona simulans*, from the Kilkieran Bay, Galway, Ireland, as previously  
3428 described (Kennedy et al., 2009). The *Streptomyces albidoflavus* J1074 strain was  
3429 provided by Dr Andriy Luzhetskyy (Helmholtz Institute for Pharmaceutical  
3430 Research Saarland, Germany). Their complete genome sequences are available  
3431 from the GenBank database (Benson et al., 2018) under the accession numbers  
3432 NZ\_CP029338 and NC\_020990, for *Streptomyces* sp. SM17 and *S. albidoflavus*  
3433 J1074, respectively. The surugamides biosynthetic gene cluster (*sur* BGC) sequence  
3434 used as a reference for this study was the one previously described in *Streptomyces*  
3435 *albidoflavus* LHW3101 (GenBank accession number: MH070261) (Zhou et al., 2019).  
3436 Other genomes used in this study's analyses were obtained from the GenBank  
3437 RefSeq database (Benson et al., 2018).

### 3438 5.2. Phylogenetic analyses

3439 The NCBI BLASTN tool (Camacho et al., 2009; Johnson et al., 2008) was used  
3440 to determine the closest 30 *Streptomyces* strains with complete genome available in  
3441 the GenBank RefSeq database (Benson et al., 2018) to the *Streptomyces* sp. SM17.  
3442 Then phylogeny analysis was performed with the concatenated sequences of the  
3443 16S rRNA, and the housekeeping genes *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB*. The  
3444 sequences were aligned using the MAFFT program (Katoh and Standley, 2013), and  
3445 the phylogeny analysis was performed using the MrBayes program (Ronquist et al.,  
3446 2012). In MrBayes, the General Time Reversible (GTR) model of nucleotide  
3447 substitution was used (Waddell and Steel, 1997), with gamma-distributed rates  
3448 across sites with a proportion of invariable sites, with 1 million generations  
3449 sampled every 100 generations. Final consensus phylogenetic tree generated by

MrBayes was processed using MEGA X (Kumar et al., 2018), with a posterior probability cut off of 95%.

Phylogeny analysis of the surugamides biosynthetic gene cluster (*sur* BGC) was performed by using the *S. albidoflavus* LHW3101 *sur* BGC nucleotide sequence as reference (Zhou et al., 2019), and searching for similar sequences on the GenBank RefSeq database using the NCBI BLASTN tool (Benson et al., 2018; Camacho et al., 2009; Johnson et al., 2008), only taking into account complete genomes. The genome regions with similarity to the *S. albidoflavus* LHW3101 *sur* BGC undergone phylogeny analysis using the same aforementioned tools and parameters.

### 5.3. Prediction of secondary metabolites biosynthetic gene clusters

In order to assess the similarities and differences between the *Streptomyces* isolates belonging to the *albidoflavus* phylogroup, in regard to their potential to produce secondary metabolites, BGCs were predicted in their genomes, using the antiSMASH (version 5) program (Blin et al., 2019). The predicted BGCs were then processed using the BiG-SCAPE program (Navarro-Muñoz et al., 2018), with the MIBiG database (version 1.4) as reference (Medema et al., 2015), and similarity clustering of gene cluster families (GCFs) was performed. The similarity network was processed using Cytoscape (Shannon et al., 2003).

### 5.4. Gene synteny analysis

The genome regions previously determined to share similarities with the *S. albidoflavus* LHW3101 *sur* BGC were manually annotated, for the known main biosynthetic genes (*surABCD*), the penicillin binding protein (PBP)-like peptide cyclase and hydrolase *surE* gene, and the gene with regulatory function *surR* (Matsuda et al., 2019a, 2019b; Ninomiya et al., 2016; Thankachan et al., 2019; Xu et al., 2017; Zhou et al., 2019). This was performed using the UniPro UGENE toolkit

(Okonechnikov et al., 2012), the GenBank database and the NCBI BLASTN tool (Benson et al., 2018; Camacho et al., 2009; Johnson et al., 2008). The gene synteny and reading frame analysis was performed using the UniPro UGENE toolkit (Okonechnikov et al., 2012) and the Artemis genome browser (Rutherford et al., 2000).

## 5.5. Diagrams and figures

All the Venn diagrams presented in this study were generated using the venn package in R (Dusa, 2018; R Core Team, 2018), and RStudio (RStudio Team, 2015). All the images presented in this study were edited using the Inkscape program (available from <https://inkscape.org/>).

## 5.6. Strains culture, maintenance, and secondary metabolites production

The same culture media and protocols were employed for both isolates *Streptomyces* sp. SM17 and *Streptomyces albidoflavus* J1074. Glycerol stocks were prepared from spores collected from soya-mannitol (SM) medium after 8 days of cultivation at 28°C and preserved at -20°C. To verify the secondary metabolites production profile, spores were cultivated for 7 days on SM agar medium at 28°C, then pre-inoculated in 5 mL TSB medium, and cultivated at 28°C and 220 rpm for 2 days. Then 10% (v/v) of the pre-inoculum was transferred to 30 mL of the following media: TSB; SYP-NaCl (1% starch, 0.4% yeast extract, 0.2% peptone & 0.1% NaCl); YD (0.4% yeast extract, 1% malt extract & 4% dextrin pH 7); P1 (2% glucose, 1% soluble starch, 0.1% meat extract, 0.4% yeast extract, 2.5% soy flour, 0.2% NaCl & 0.005% K<sub>2</sub>HPO<sub>4</sub> pH 7.3); P2 (1 % glucose, 0.6% glycerol, 0.1% yeast extract, 0.2% malt extract, 0.6% MgCl<sub>2</sub>·6H<sub>2</sub>O, 0.03% CaCO<sub>3</sub> & 10% sea water); P3 (2.5% soy flour, 0.75% starch, 2.25% glucose, 0.35% yeast extract, 0.05% ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.6% CaCO<sub>3</sub> pH 6); CH-F2 (2 % soy flour, 0.5 % yeast extract, 0.2 % CaCO<sub>3</sub>, 0.05 % citric acid, 5 % glucose, pH 7.0); SY (2.5% soluble starch, 1.5% soy flour, 0.2% yeast extract &

0.4% CaCO<sub>3</sub> pH7); Sporulation medium (2% soluble starch & 0.4 yeast extract); and Oatmeal medium (2% oatmeal). These were cultivated at 28°C and 220 rpm for 4 days in TSB; and for 8 days in SYP-NaCl, YD, SY, P1, P2, P3, CH-F2, Sporulation, and Oatmeal media. Once the bioprocess was completed, the broth was frozen at -20°C for further chemical analysis.

### 5.7. Metabolic profiling, compound isolation and chemical structure analysis

The *Streptomyces* broth of TSB, SYP-NaCl and YD medium cultures (180 mL) was exhaustively extracted using a solvent mixture of 1:1 MeOH:DCM yielding a crude extract (3.89 g). This crude extract was first separated using SPE on C<sub>18</sub> bonded silica gel (Polygoprep C18, 12%C, 60Å, 40-63 µm), eluting with varying solvent mixtures to produce five fractions: H<sub>2</sub>O (743.62 mg), 1:1 H<sub>2</sub>O:MeOH (368.6 mg), MeOH (15.4 mg), 1:1 MeOH:DCM (10.9 mg), DCM (8.2 mg). The final three fractions (MeOH, 1:1 MeOH:DCM, DCM, 34.5 mg) were then combined and subject to analytical reverse phase HPLC on a Waters Symmetry C18 5 µm, 4.6 x 250 mm column. The column was eluted with 10% MeCN (0.1% TFA)/90% H<sub>2</sub>O (0.1% TFA) for 5 min, then a linear gradient to 100% MeCN (0.1% TFA) over 21 min was performed. The column was further eluted with 100% MeCN (0.1% TFA) for 6 min. After the HPLC was complete a linear gradient back to 10% MeCN (0.1% TFA)/90% H<sub>2</sub>O (0.1% TFA) over 1 min and then further elution of 10% MeCN (0.1% FA)/90% H<sub>2</sub>O (0.1% FA) for 4 min was performed. This yielded pure surugamide A (0.8 mg). Surugamide A was characterised using MS and NMR data to confirm the structure for use as an analytical standard.

Surugamide A was quantified in the broth using LC-MS analysis on an Agilent UHR-qTOF 6540 mass spectrometer. The column used for separation was Waters equity UPLC BEH C18 1.7 µm 2.1 x 75 mm. The column was eluted with 10% MeCN (0.1% FA)/90% H<sub>2</sub>O (0.1% FA) for 2 min, then a linear gradient to 100%

3528 MeCN (0.1% FA) over 6 min was performed. The column was further eluted with  
3529 100% MeCN (0.1% FA) for 4 min. After the UPLC was complete a linear gradient  
3530 back to 10% MeCN (0.1% FA)/90% H<sub>2</sub>O (0.1% FA) over 1 min and then further  
3531 elution of 10% MeCN (0.1% FA)/90% H<sub>2</sub>O (0.1% FA) for 3 min was performed  
3532 before the next run. The MS detection method was positive ion. A calibration curve  
3533 was produce using the LC-MS method above and injecting the pure surugamide A  
3534 at seven concentrations (100, 25, 10, 2, 1, 0.2, 0.1 mg/L). 30 mL of each *Streptomyces*  
3535 strain in broth were extracted using a solvent mixture of 1:1 MeOH:DCM three  
3536 times to yield a crude extract. These extracts were resuspended in MeOH and  
3537 filtered through PTFE 0.2 µm filters before being subject to the above LC-MS  
3538 method.

3539       The surugamide A calibration standards 1-7 and the six extracts were  
3540 analysed using the Agilent MassHunter Quantification software package. This  
3541 allowed the quantification of surugamide A in the extracts based on the intensity of  
3542 peaks in the chromatogram with matching retention time and exact mass.



## 3543 6. References

- 3544 Almeida, E. L. de, Carillo Rincon, F., Jackson, S. A., and Dobson, A. D. (2019).  
 3545 Comparative genomics of marine sponge-derived *Streptomyces* spp. isolates  
 3546 SM17 and SM18 with their closest terrestrial relatives provides novel insights  
 3547 into environmental niche adaptations and secondary metabolite biosynthesis  
 3548 potential. *Front. Microbiol.* 10, 1713. doi:10.3389/FMICB.2019.01713.
- 3549 Almeida, E. L., Margassery, L. M., Kennedy, J., and Dobson, A. D. W. (2018). Draft  
 3550 Genome Sequence of the Antimycin-Producing Bacterium *Streptomyces* sp.  
 3551 Strain SM8, Isolated from the Marine Sponge *Haliclona simulans*. *Genome*  
 3552 *Announc.* 6, e01535-17. doi:10.1128/genomeA.01535-17.
- 3553 Badalamenti, J. P., Erickson, J. D., and Salomon, C. E. (2016). Complete Genome  
 3554 Sequence of *Streptomyces albus* SM254, a Potent Antagonist of Bat White-Nose  
 3555 Syndrome Pathogen *Pseudogymnoascus destructans*. *Genome Announc.* 4.  
 3556 doi:10.1128/genomeA.00290-16.
- 3557 Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D.,  
 3558 et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47.  
 3559 doi:10.1093/nar/gkx1094.
- 3560 Bilyk, O., Sekurova, O. N., Zotchev, S. B., and Luzhetskyy, A. (2016). Cloning and  
 3561 Heterologous Expression of the Grecoacycline Biosynthetic Gene Cluster. *PLoS*  
 3562 *One* 11, e0158682. doi:10.1371/journal.pone.0158682.
- 3563 Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019).  
 3564 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.  
 3565 *Nucleic Acids Res.* doi:10.1093/nar/gkz310.
- 3566 Bose, U., Hewavitharana, A., Ng, Y., Shaw, P., Fuerst, J., and Hodson, M. (2015). LC-  
 3567 MS-Based Metabolomics Study of Marine Bacterial Secondary Metabolite and  
 3568 Antibiotic Production in *Salinispora arenicola*. *Mar. Drugs* 13, 249–266.  
 3569 doi:10.3390/md13010249.
- 3570 Brückner, R., and Titgemeyer, F. (2002). Carbon catabolite repression in bacteria:  
 3571 Choice of the carbon source and autoregulatory limitation of sugar utilization.  
 3572 *FEMS Microbiol. Lett.* 209, 141–148. doi:10.1016/S0378-1097(02)00559-1.
- 3573 Calcabrini, C., Catanzaro, E., Bishayee, A., Turrini, E., and Fimognari, C. (2017).  
 3574 Marine Sponge Natural Products with Anticancer Potential: An Updated  
 3575 Review. *Mar. Drugs* 15, 310. doi:10.3390/md15100310.
- 3576 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al.  
 3577 (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.  
 3578 doi:10.1186/1471-2105-10-421.

- 3579 Chater, K. F., and Wilde, L. C. (1976). Restriction of a bacteriophage of  
3580 *Streptomyces albus* G involving endonuclease *SaI*. *J. Bacteriol.* 128, 644–50.
- 3581 Chater, K. F., and Wilde, L. C. (1980). *Streptomyces albus* G Mutants Defective in  
3582 the *SaI*GI Restriction-Modification System. *Microbiology* 116, 323–334.  
3583 doi:10.1099/00221287-116-2-323.
- 3584 Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., et al. (2003).  
3585 Organizational and Mutational Analysis of a Complete FR-008/Candididin  
3586 Gene Cluster Encoding a Structurally Related Polyene Complex. *Chem. Biol.*  
3587 10, 1065–1076. doi:10.1016/J.CHEMBIOL.2003.10.007.
- 3588 Cuthbertson, L., and Nodwell, J. R. (2013). The TetR Family of Regulators.  
3589 *Microbiol. Mol. Biol. Rev.* 77, 440–475. doi:10.1128/MMBR.00018-13.
- 3590 Demers, D., Knestrick, M., Fleeman, R., Tawfik, R., Azhari, A., Souza, A., et al.  
3591 (2018). Exploitation of Mangrove Endophytic Fungi for Infectious Disease  
3592 Drug Discovery. *Mar. Drugs* 16, 376. doi:10.3390/md16100376.
- 3593 Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria.  
3594 *Curr. Opin. Microbiol.* 11, 87–93. doi:10.1016/j.mib.2008.02.007.
- 3595 Dusa, A. (2018). venn: Draw Venn Diagrams. Available at: [https://cran.r-](https://cran.r-project.org/package=venn)  
3596 [project.org/package=venn](https://cran.r-project.org/package=venn).
- 3597 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al.  
3598 (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47,  
3599 D427–D432. doi:10.1093/nar/gky995.
- 3600 Gadagkar, S. R., Rosenberg, M. S., and Kumar, S. (2005). Inferring species  
3601 phylogenies from multiple genes: Concatenated sequence tree versus  
3602 consensus gene tree. *J. Exp. Zool. Part B Mol. Dev. Evol.* 304B, 64–74.  
3603 doi:10.1002/jez.b.21026.
- 3604 Gallo, M., and Katz, E. (1972). Regulation of secondary metabolite biosynthesis:  
3605 catabolite repression of phenoxazinone synthase and actinomycin formation  
3606 by glucose. *J. Bacteriol.* 109, 659–67.
- 3607 Glaeser, S. P., and Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in  
3608 prokaryotic taxonomy. *Syst. Appl. Microbiol.* 38, 237–245.  
3609 doi:10.1016/j.syapm.2015.03.007.
- 3610 Glavinas, H., Krajcsi, P., Cserepes, J., and Sarkadi, B. (2004). The role of ABC  
3611 transporters in drug resistance, metabolism and toxicity. *Curr. Drug Deliv.* 1,  
3612 27–42.
- 3613 Hamm, P. S., Caimi, N. A., Northup, D. E., Valdez, E. W., Buecher, D. C., Dunlap, C.  
3614 A., et al. (2017). Western Bats as a Reservoir of Novel *Streptomyces* Species

- 3615 with Antifungal Activity. *Appl. Environ. Microbiol.* 83, e03057-16.  
3616 doi:10.1128/AEM.03057-16.
- 3617 Hoz, J. F.-D. la, Méndez, C., Salas, J. A., and Olano, C. (2017). Novel Bioactive  
3618 Paulomycin Derivatives Produced by *Streptomyces albus* J1074. *Molecules* 22,  
3619 1758. doi:10.3390/molecules22101758.
- 3620 Huang, C., Yang, C., Zhang, W., Zhu, Y., Ma, L., Fang, Z., et al. (2019). Albumycin, a  
3621 new isoindolequinone from *Streptomyces albus* J1074 harboring the fluostatin  
3622 biosynthetic gene cluster. *J. Antibiot. (Tokyo)*. 72, 311–315. doi:10.1038/s41429-  
3623 019-0161-4.
- 3624 Hwang, K.-S., Kim, H. U., Charusanti, P., Palsson, B. Ø., and Lee, S. Y. (2014).  
3625 Systems biology and biotechnology of *Streptomyces* species for the production  
3626 of secondary metabolites. *Biotechnol. Adv.* 32, 255–268.  
3627 doi:10.1016/j.biotechadv.2013.10.008.
- 3628 Indraningrat, A., Smidt, H., and Sipkema, D. (2016). Bioprospecting Sponge-  
3629 Associated Microbes for Antimicrobial Compounds. *Mar. Drugs* 14, 87.  
3630 doi:10.3390/md14050087.
- 3631 Inoue, O. O., Schmidell Netto, W., Padilla, G., and Facciotti, M. C. R. (2007). Carbon  
3632 catabolite repression of retamycin production by *Streptomyces olindensis*  
3633 ICB20. *Brazilian J. Microbiol.* 38, 58–61. doi:10.1590/S1517-83822007000100012.
- 3634 Jackson, S., Crossman, L., Almeida, E., Margassery, L., Kennedy, J., and Dobson, A.  
3635 (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene  
3636 Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates.  
3637 *Mar. Drugs* 16, 67. doi:10.3390/md16020067.
- 3638 Jiang, G., Zhang, Y., Powell, M. M., Zhang, P., Zuo, R., Zhang, Y., et al. (2018). High-  
3639 Yield Production of Herbicidal Thaxtomins and Thaxtomin Analogs in a  
3640 Nonpathogenic *Streptomyces* Strain. *Appl. Environ. Microbiol.* 84, e00164-18.  
3641 doi:10.1128/AEM.00164-18.
- 3642 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden,  
3643 T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–  
3644 W9. doi:10.1093/nar/gkn201.
- 3645 Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment  
3646 Software Version 7: Improvements in Performance and Usability. *Mol. Biol.*  
3647 *Evol.* 30, 772–780. doi:10.1093/molbev/mst010.
- 3648 Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., et al. (2009).  
3649 Isolation and Analysis of Bacteria with Antimicrobial Activities from the  
3650 Marine Sponge *Haliclona simulans* Collected from Irish Waters. *Mar.*  
3651 *Biotechnol.* 11, 384–396. doi:10.1007/s10126-008-9154-1.

- 3652 Koshla, O. T., Rokytskyy, I. V., Ostash, I. S., Busche, T., Kalinowski, J., Mösker, E., et  
3653 al. (2019). Secondary Metabolome and Transcriptome of *Streptomyces albus*  
3654 J1074 in Liquid Medium SG2. *Cytol. Genet.* 53, 1–7.  
3655 doi:10.3103/S0095452719010080.
- 3656 Koskiniemi, S., Lamoureux, J. G., Nikolakakis, K. C., t’Kint de Roodenbeke, C.,  
3657 Kaplan, M. D., Low, D. A., et al. (2013). Rhs proteins from diverse bacteria  
3658 mediate intercellular competition. *Proc. Natl. Acad. Sci.* 110, 7032–7037.  
3659 doi:10.1073/pnas.1300627110.
- 3660 Kremling, A., Geiselmann, J., Ropers, D., and de Jong, H. (2015). Understanding  
3661 carbon catabolite repression in *Escherichia coli* using quantitative models.  
3662 *Trends Microbiol.* 23, 99–109. doi:10.1016/j.tim.2014.11.002.
- 3663 Kumar, S., He, G., Kakarla, P., Shrestha, U., Ranjana, K. C., Ranaweera, I., et al.  
3664 (2016). Bacterial Multidrug Efflux Pumps of the Major Facilitator Superfamily  
3665 as Targets for Modulation. *Infect. Disord. Drug Targets* 16, 28–43.
- 3666 Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X:  
3667 Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol.*  
3668 *Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.
- 3669 Kunz, A., Labes, A., Wiese, J., Bruhn, T., Bringmann, G., and Imhoff, J. (2014).  
3670 Nature’s Lab for Derivatization: New and Revised Structures of a Variety of  
3671 Streptophenazines Produced by a Sponge-Derived *Streptomyces* Strain. *Mar.*  
3672 *Drugs* 12, 1699–1714. doi:10.3390/md12041699.
- 3673 Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., et al.  
3674 (2018). Total Synthesis of the Nonribosomal Peptide Surugamide B and  
3675 Identification of a New Offloading Cyclase Family. *Angew. Chemie Int. Ed.* 57,  
3676 9447–9451. doi:10.1002/anie.201805541.
- 3677 Labeda, D. P., Doroghazi, J. R., Ju, K.-S., and Metcalf, W. W. (2014). Taxonomic  
3678 evaluation of *Streptomyces albus* and related species using multilocus  
3679 sequence analysis and proposals to emend the description of *Streptomyces*  
3680 *albus* and describe *Streptomyces pathocidini* sp. nov. *Int. J. Syst. Evol.*  
3681 *Microbiol.* 64, 894–900. doi:10.1099/ij.s.0.058107-0.
- 3682 Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K.-S., et al.  
3683 (2017). Phylogenetic relationships in the family Streptomycetaceae using  
3684 multi-locus sequence analysis. *Antonie Van Leeuwenhoek* 110, 563–583.  
3685 doi:10.1007/s10482-016-0824-0.
- 3686 Lautru, S., Oves-Costales, D., Pernodet, J.-L., and Challis, G. L. (2007). MbtH-like  
3687 protein-mediated cross-talk between non-ribosomal peptide antibiotic and

- 3688 siderophore biosynthetic pathways in *Streptomyces coelicolor* M145.  
3689 *Microbiology* 153, 1405–1412. doi:10.1099/mic.0.2006/003145-0.
- 3690 Li, J., Xie, Z., Wang, M., Ai, G., and Chen, Y. (2015). Identification and Analysis of  
3691 the Paulomycin Biosynthetic Gene Cluster and Titer Improvement of the  
3692 Paulomycins in *Streptomyces paulus* NRRL 8115. *PLoS One* 10, e0120542.  
3693 doi:10.1371/journal.pone.0120542.
- 3694 Li, S., Zhang, B., Zhu, H., and Zhu, T. (2018). Cloning and Expression of the  
3695 Chitinase Encoded by *ChiKJ406136* from *Streptomyces sampsonii* (Millard &  
3696 Burr) Waksman KJ40 and Its Antifungal Effect. *Forests* 9, 699.  
3697 doi:10.3390/f9110699.
- 3698 Li, Y., Pinto-Tomás, A. A., Rong, X., Cheng, K., Liu, M., and Huang, Y. (2019).  
3699 Population Genomics Insights into Adaptive Evolution and Ecological  
3700 Differentiation in Streptomycetes. *Appl. Environ. Microbiol.* 85.  
3701 doi:10.1128/AEM.02555-18.
- 3702 Liu, Q., Xiao, L., Zhou, Y., Deng, K., Tan, G., Han, Y., et al. (2016). Development of  
3703 *Streptomyces* sp. FR-008 as an emerging chassis. *Synth. Syst. Biotechnol.* 1,  
3704 207–214. doi:10.1016/J.SYNBIO.2016.07.002.
- 3705 Magnus, N., Weise, T., and Piechulla, B. (2017). Carbon Catabolite Repression  
3706 Regulates the Production of the Unique Volatile Sodorifen of *Serratia*  
3707 *plymuthica* 4Rx13. *Front. Microbiol.* 8, 2522. doi:10.3389/fmicb.2017.02522.
- 3708 Manteca, A., Alvarez, R., Salazar, N., Yague, P., and Sanchez, J. (2008). Mycelium  
3709 Differentiation and Antibiotic Production in Submerged Cultures of  
3710 *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* 74, 3877–3886.  
3711 doi:10.1128/AEM.02715-07.
- 3712 Manteca, Á., and Yagüe, P. (2019). “*Streptomyces* as a Source of Antimicrobials:  
3713 Novel Approaches to Activate Cryptic Secondary Metabolite Pathways,” in  
3714 *Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity*  
3715 *Methods* (IntechOpen). doi:10.5772/intechopen.81812.
- 3716 Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L.  
3717 Y., et al. (2015). CDD: NCBI’s conserved domain database. *Nucleic Acids Res.*  
3718 43, D222–D226. doi:10.1093/nar/gku1221.
- 3719 Matsuda, K., Kobayashi, M., Kuranaga, T., Takada, K., Ikeda, H., Matsunaga, S., et  
3720 al. (2019a). SurE is a trans -acting thioesterase cyclizing two distinct non-  
3721 ribosomal peptides. *Org. Biomol. Chem.* 17, 1058–1061.  
3722 doi:10.1039/C8OB02867B.

- 3723 Matsuda, K., Kuranaga, T., Sano, A., Ninomiya, A., Takada, K., and Wakimoto, T.  
 3724 (2019b). The Revised Structure of the Cyclic Octapeptide Surugamide A. *Chem.*  
 3725 *Pharm. Bull.* 67, 476–480. doi:10.1248/cpb.c19-00002.
- 3726 Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et  
 3727 al. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem.*  
 3728 *Biol.* 11, 625–631. doi:10.1038/nchembio.1890.
- 3729 Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., and  
 3730 Luzhetskyy, A. (2018). Generation of a cluster-free *Streptomyces albus* chassis  
 3731 strains for improved heterologous expression of secondary metabolite clusters.  
 3732 *Metab. Eng.* 49, 316–324. doi:10.1016/j.ymben.2018.09.004.
- 3733 Navarro-Muñoz, J. C., Selem-Mojica, N., Mullaney, M. W., Kautsar, S., Tryon, J. H.,  
 3734 Parkinson, E. I., et al. (2018). A computational framework for systematic  
 3735 exploration of biosynthetic diversity from large-scale genomic data. *bioRxiv*,  
 3736 445270. doi:10.1101/445270.
- 3737 Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., et al.  
 3738 (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an  
 3739 Unrelated Decapeptide, Surugamide F. *ChemBioChem* 17, 1709–1712.  
 3740 doi:10.1002/cbic.201600350.
- 3741 Ohki, R., Tatenno, K., Okada, Y., Okajima, H., Asai, K., Sadaie, Y., et al. (2003). A  
 3742 Bacitracin-Resistant *Bacillus subtilis* Gene Encodes a Homologue of the  
 3743 Membrane-Spanning Subunit of the *Bacillus licheniformis* ABC Transporter. *J.*  
 3744 *Bacteriol.* 185, 51–59. doi:10.1128/JB.185.1.51-59.2003.
- 3745 Okonechnikov, K., Golosova, O., and Fursov, M. (2012). Unipro UGENE: a unified  
 3746 bioinformatics toolkit. *Bioinformatics* 28, 1166–1167.  
 3747 doi:10.1093/bioinformatics/bts091.
- 3748 Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et  
 3749 al. (2015). Roary: rapid large-scale prokaryote pan genome analysis.  
 3750 *Bioinformatics* 31, 3691–3693. doi:10.1093/bioinformatics/btv421.
- 3751 Pan, R., Bai, X., Chen, J., Zhang, H., and Wang, H. (2019). Exploring Structural  
 3752 Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A  
 3753 Literature Review. *Front. Microbiol.* 10, 294. doi:10.3389/fmicb.2019.00294.
- 3754 Podlesek, Z., Comino, A., Herzog-Velikonja, B., Zgur-Bertok, D., Komel, R., and  
 3755 Grabnar, M. (1995). *Bacillus licheniformis* bacitracin-resistance ABC  
 3756 transporter: relationship to mammalian multidrug resistance. *Mol. Microbiol.*  
 3757 16, 969–76.
- 3758 Polgar, O., and Bates, S. E. (2005). ABC transporters in the balance: is there a role in  
 3759 multidrug resistance? *Biochem. Soc. Trans.* 33, 241–5. doi:10.1042/BST0330241.

- 3760 Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H., and Walsh, C. T. (1998).  
 3761 Identification of a *Mycobacterium tuberculosis* gene cluster encoding the  
 3762 biosynthetic enzymes for assembly of the virulence-conferring siderophore  
 3763 mycobactin. *Chem. Biol.* 5, 631–45.
- 3764 R Core Team (2018). R: A Language and Environment for Statistical Computing.
- 3765 Romano, S., Jackson, S., Patry, S., and Dobson, A. (2018). Extending the “One Strain  
 3766 Many Compounds” (OSMAC) Principle to Marine Microorganisms. *Mar.*  
 3767 *Drugs* 16, 244. doi:10.3390/md16070244.
- 3768 Romero-Rodríguez, A., Ruiz-Villafán, B., Tierrafría, V. H., Rodríguez-Sanoja, R.,  
 3769 and Sánchez, S. (2016). Carbon Catabolite Regulation of Secondary Metabolite  
 3770 Formation and Morphological Differentiation in *Streptomyces coelicolor*. *Appl.*  
 3771 *Biochem. Biotechnol.* 180, 1152–1166. doi:10.1007/s12010-016-2158-9.
- 3772 Rong, X., Doroghazi, J. R., Cheng, K., Zhang, L., Buckley, D. H., and Huang, Y.  
 3773 (2013). Classification of *Streptomyces* phylogroup *pratensis* (Doroghazi and  
 3774 Buckley, 2010) based on genetic and phenotypic evidence, and proposal of  
 3775 *Streptomyces pratensis* sp. nov. *Syst. Appl. Microbiol.* 36, 401–407.  
 3776 doi:10.1016/j.syapm.2013.03.010.
- 3777 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et  
 3778 al. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model  
 3779 Choice Across a Large Model Space. *Syst. Biol.* 61, 539–542.  
 3780 doi:10.1093/sysbio/sys029.
- 3781 RStudio Team (2015). RStudio: Integrated Development Environment for R.  
 3782 Available at: <http://www.rstudio.com/>.
- 3783 Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al.  
 3784 (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–  
 3785 945. doi:10.1093/bioinformatics/16.10.944.
- 3786 Sathiyarayanan, G., Gandhimathi, R., Sabarathnam, B., Seghal Kiran, G., and  
 3787 Selvin, J. (2014). Optimization and production of pyrrolidone antimicrobial  
 3788 agent from marine sponge-associated *Streptomyces* sp. MAPS15. *Bioprocess*  
 3789 *Biosyst. Eng.* 37, 561–573. doi:10.1007/s00449-013-1023-2.
- 3790 Schneemann, I., Kajahn, I., Ohlendorf, B., Zinecker, H., Erhard, A., Nagel, K., et al.  
 3791 (2010). Mayamycin, a Cytotoxic Polyketide from a *Streptomyces* Strain Isolated  
 3792 from the Marine Sponge *Halichondria panicea*. *J. Nat. Prod.* 73, 1309–1312.  
 3793 doi:10.1021/np100135b.
- 3794 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al.  
 3795 (2003). Cytoscape: a software environment for integrated models of

- 3796 biomolecular interaction networks. *Genome Res.* 13, 2498–504.  
3797 doi:10.1101/gr.1239303.
- 3798 Stülke, J., and Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr.*  
3799 *Opin. Microbiol.* 2, 195–201. doi:10.1016/S1369-5274(99)80034-4.
- 3800 Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., et al. (2013).  
3801 Surugamides A–E, Cyclic Octapeptides with Four d-Amino Acid Residues,  
3802 from a Marine *Streptomyces* sp.: LC–MS-Aided Inspection of Partial  
3803 Hydrolysates for the Distinction of d- and l-Amino Acid Residues in the  
3804 Sequence. *J. Org. Chem.* 78, 6746–6750. doi:10.1021/jo400708u.
- 3805 Thabit, A. K., Crandon, J. L., and Nicolau, D. P. (2015). Antimicrobial resistance:  
3806 impact on clinical and economic outcomes and the need for new  
3807 antimicrobials. *Expert Opin. Pharmacother.* 16, 159–177.  
3808 doi:10.1517/14656566.2015.993381.
- 3809 Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M. E., and Seipke, R. F.  
3810 (2019). A *trans*-Acting Cyclase Offloading Strategy for Nonribosomal Peptide  
3811 Synthetases. *ACS Chem. Biol.* 14, 845–849. doi:10.1021/acscchembio.9b00095.
- 3812 Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I., and Miller, A. A. (2015).  
3813 ESKAPEing the labyrinth of antibacterial discovery. *Nat. Rev. Drug Discov.* 14,  
3814 529–542. doi:10.1038/nrd4572.
- 3815 Viegelmann, C., Margassery, L., Kennedy, J., Zhang, T., O'Brien, C., O'Gara, F., et al.  
3816 (2014). Metabolomic Profiling and Genomic Study of a Marine Sponge-  
3817 Associated *Streptomyces* sp. *Mar. Drugs* 12, 3323–3351.  
3818 doi:10.3390/md12063323.
- 3819 Waddell, P. J., and Steel, M. . (1997). General Time-Reversible Distances with  
3820 Unequal Rates across Sites: Mixing  $\Gamma$  and Inverse Gaussian Distributions with  
3821 Invariant Sites. *Mol. Phylogenet. Evol.* 8, 398–414. doi:10.1006/mpev.1997.0452.
- 3822 Ward, A., and Allenby, N. (2018). Genome mining for the search and discovery of  
3823 bioactive compounds: The *Streptomyces* paradigm. *FEMS Microbiol. Lett.* 365.  
3824 doi:10.1093/femsle/fny240.
- 3825 Xu, F., Nazari, B., Moon, K., Bushin, L. B., and Seyedsayamdost, M. R. (2017).  
3826 Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074  
3827 Using High-Throughput Elicitor Screens. *J. Am. Chem. Soc.* 139, 9203–9212.  
3828 doi:10.1021/jacs.7b02716.
- 3829 Yan, N. (2015). Structural Biology of the Major Facilitator Superfamily Transporters.  
3830 *Annu. Rev. Biophys.* 44, 257–283. doi:10.1146/annurev-biophys-060414-033901.

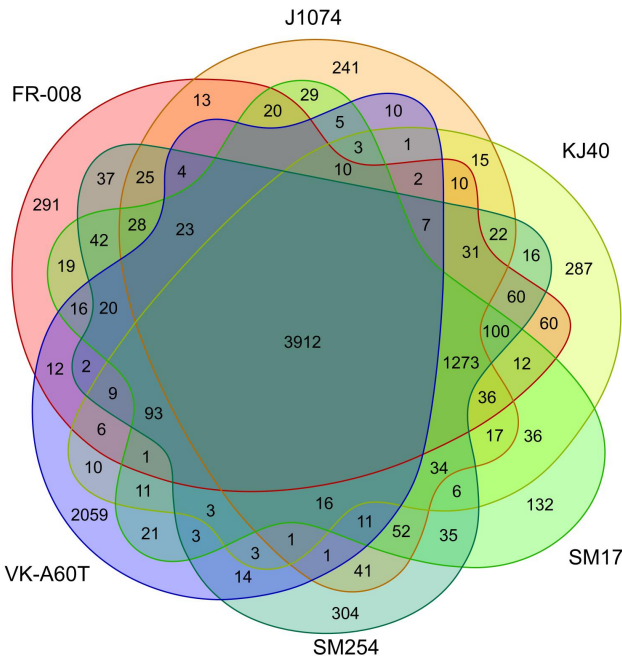


- 3831 Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., and Luzhetskyy, A. (2014).  
3832 Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC*  
3833 *Genomics* 15, 97. doi:10.1186/1471-2164-15-97.
- 3834 Zhao, Z., Deng, Z., Pang, X., Chen, X.-L., Zhang, P., Bai, L., et al. (2015). Production  
3835 of the antibiotic FR-008/candicidin in *Streptomyces* sp. FR-008 is co-regulated  
3836 by two regulators, FscRI and FscRIV, from different transcription factor  
3837 families. *Microbiology* 161, 539–552. doi:10.1099/mic.0.000033.
- 3838 Zhou, Y., Lin, X., Xu, C., Shen, Y., Wang, S.-P., Liao, H., et al. (2019). Investigation of  
3839 Penicillin Binding Protein (PBP)-like Peptide Cyclase and Hydrolase in  
3840 Surugamide Non-ribosomal Peptide Biosynthesis. *Cell Chem. Biol.* 26, 737-  
3841 744.e4. doi:10.1016/j.chembiol.2019.02.010.

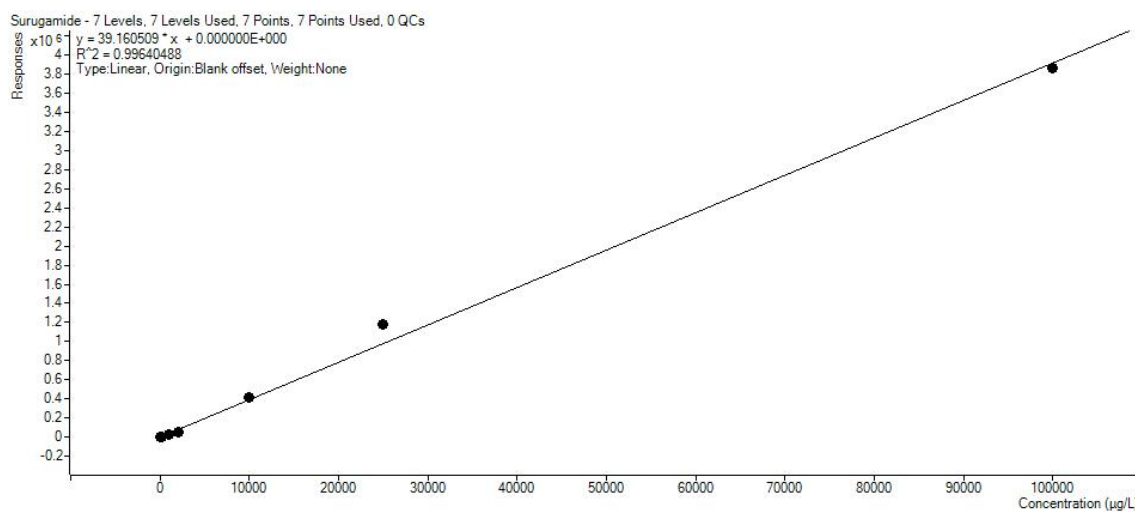
3842 7. Supplementary material

Isolate	Number of bases	Number of CDSs	rRNA	tRNA	tmRNA
FR-008	7,090,955	6,126	21	79	1
J1074	6,841,649	5,847	21	77	1
KJ40	7,070,328	6,057	21	78	1
SM17	6,975,788	5,972	21	78	1
SM254	7,170,504	6,182	21	77	1

3843 **Table S1:** Genome statistics determined using the Prokka program, of the  
3844 *Streptomyces* isolates genomes obtained from GenBank and determined to belong  
3845 to the *albidoflavus* phylogroup.



3846 **Figure S1:** Venn diagram representing the presence/absence of groups of  
3847 orthologous genes in the *albidoflavus* phylogroup genomes (namely strains FR-008,  
3848 J1074, KJ40, SM17, and SM254), also including the *Streptomyces koyangensis* VK-  
3849 A60T genome.



3850 **Figure S2:** Calibration curve for surugamide A, determined using LC-MS and pure  
 3851 surugamide A at seven concentrations (0.1, 0.2, 1, 2, 10, 25, 100 mg/L).

# Chapter 4

*In silico* screening and heterologous expression of a Polyethylene Terephthalate hydrolase (PETase)-like enzyme (SM14est) with Polycaprolactone (PCL)-degrading activity, from the marine sponge-derived strain *Streptomyces* sp. SM14

Eduardo L. Almeida<sup>1</sup>, Andrés Felipe Carrillo Rincón<sup>1</sup>, Stephen A. Jackson<sup>1,2</sup> and Alan D.W. Dobson<sup>1,2</sup>

<sup>1</sup>School of Microbiology, University College Cork, Cork, Ireland

<sup>2</sup>Environmental Research Institute, University College Cork, Cork, Ireland

The work presented in this chapter has been published in the journal *Frontiers in Microbiology*, 01 October 2019, <https://doi.org/10.3389/fmicb.2019.02187>.

**Citation:** Almeida E. L., Carrillo Rincón A. F., Jackson S. A. and Dobson A. D. W. (2019). *In silico* Screening and Heterologous Expression of a Polyethylene Terephthalate Hydrolase (PETase)-Like Enzyme (SM14est) With Polycaprolactone (PCL)-Degrading Activity, From the Marine Sponge-Derived Strain *Streptomyces* sp. SM14. *Front. Microbiol.* 10:2187. doi: 10.3389/fmicb.2019.02187

3852    **1.    Abstract**

3853           Plastics, such as the polyethylene terephthalate (PET), are widely used for various  
3854 industrial applications, due to their physicochemical properties which are particularly  
3855 useful in the packaging industry. However, due to improper plastic waste management  
3856 and difficulties in recycling; post-consumer plastic waste has become a pressing issue for  
3857 both the environment and for human health. Hence, novel technologies and methods of  
3858 processing plastic waste are required to address these issues. Enzymatic-assisted  
3859 hydrolysis of synthetic polymers has been proposed as a potentially more efficient and  
3860 environment-friendly alternative to the currently employed methods. Recently, a number  
3861 of PET hydrolases have been described, and in particular a PETase derived from the  
3862 *Ideonella sakaiensis* 201-F6 (IsPETase), which appears to be the most efficient and  
3863 substrate-specific bacterial PET hydrolase enzyme discovered to date. In order to further  
3864 investigate the class of PETase-like enzymes, we employed an *in silico*-based screening  
3865 approach on the biotechnologically relevant genus *Streptomyces*, including terrestrial  
3866 and marine isolates; in a search for potential PETase homologs. From a total of 52  
3867 genomes analysed, we were able to identify 3 potential PETase-like enzymes, all of which  
3868 were derived from marine-sponge associated *Streptomyces* isolates. A candidate PETase-  
3869 like gene (SM14est) was identified in *Streptomyces* sp. SM14. Further *in silico*  
3870 characterisation of the SM14est protein sequence and its potential three-dimensional  
3871 structure were performed and compared to the well characterised IsPETase. Both the  
3872 serine hydrolase motif Gly-x1-Ser-x2-Gly and the catalytic triad Ser, Asp, His are  
3873 conserved in both sequences. Molecular docking experiments indicated that the SM14est  
3874 enzyme possessed the capacity to bind plastics as substrates. Finally, polyesterase activity  
3875 was confirmed using a polycaprolactone (PCL) plate clearing assay which is a model  
3876 substrate for plastics degradation; following heterologous expression of SM14est in *E.*  
3877 *coli*, with secretion being facilitated by the native *Streptomyces* signal peptide. These  
3878 findings provide further insights into this important class of PETase-like enzymes.

## 3879 2. Introduction

3880           Plastics are materials that have been produced on a large scale from the  
3881 1950s onwards, and since then have been widely used for various applications, and  
3882 have become almost indispensable in modern society (Geyer et al., 2017; Jambeck et  
3883 al., 2015; Lebreton and Andrady, 2019). In the 1960s, plastics accounted for less than  
3884 1% of municipal solid waste in the United States, but steadily increased to around  
3885 10% by 2005 in countries with middle to high income (Geyer et al., 2017; Jambeck et  
3886 al., 2015). This was largely due to their advantageous properties, such as their low  
3887 production cost and bio-inertia, which are particularly useful for the packaging  
3888 industry, when compared to other materials. This has resulted in the use of plastics  
3889 in the packaging sector, which accounts for around 40% of the plastic converter  
3890 demand in Europe (Lebreton and Andrady, 2019; PlasticsEurope, 2018). However,  
3891 some of these aforementioned characteristics have resulted in plastics becoming a  
3892 critical problem from an environmental perspective; as many synthetic plastics are  
3893 highly recalcitrant to biodegradation and can persist for long periods of time in the  
3894 environment (Wei and Zimmermann, 2017a).

3895           In 2017, there was an estimated worldwide plastics production of 348  
3896 million tonnes – an increase from the 335 million tonnes estimated for the previous  
3897 year, and this does not include polyethylene terephthalate (PET)-, polyamide (PA)-  
3898 and polyacryl-fibres) (PlasticsEurope, 2018). At the current rate, this number is  
3899 expected to double in the next 20 years (Lebreton and Andrady, 2019). In the past  
3900 decade, plastic waste management policies have helped considerably in reducing  
3901 post-consumer plastic waste being disposed in the environment. For example, in  
3902 Europe more collected plastic waste (31.1% of 27.1 million tonnes of collected  
3903 plastic waste in 2016) was submitted to recycling rather than to landfills, for the  
3904 first time. However, landfills and incineration for energy recovery still account for

3905 27,3% and 41.6% of the collected plastic post-consumer waste, respectively  
3906 (PlasticsEurope, 2018).

3907         Notwithstanding this, the aforementioned metrics do not take into account  
3908 global mismanaged plastic waste which enters the natural environment, at  
3909 locations others than landfills. A recent study has estimated that between 60 and 99  
3910 million metric tonnes of mismanaged plastic waste was produced globally in 2015,  
3911 and that this number could triple by 2060 (Lebreton and Andrady, 2019).  
3912 Mismanaged plastic waste is particularly concerning when it effects the marine  
3913 environment. It has been calculated that in 2010 between 4.8 and 12.7 million  
3914 metric tons of plastic waste entered our oceans, and this data only accounted for  
3915 coastal countries (Jambeck et al., 2015). Sunlight and other weathering effects cause  
3916 the fragmentation of plastic debris into milli- and micro-metric particles (< 5mm),  
3917 which are defined as micro-plastics (Geyer et al., 2017; Lebreton and Andrady,  
3918 2019). These micro-plastics are now believed to be ubiquitous in soil and aquatic  
3919 environments, and are commonly ingested by animals (Lebreton and Andrady,  
3920 2019; Santillo et al., 2017). This is especially concerning, since micro-plastics can  
3921 absorb and concentrate pollutants present in the ocean and transfer them along the  
3922 food chain, particularly to seafood species that are consumed by humans (Santillo  
3923 et al., 2017). Highlighting the issue of the ubiquitous presence of micro-plastics in  
3924 the marine environment, a recent study detected the ingestion of micro-plastics by  
3925 deep-sea amphipods, at depths ranging from 7,000 to 10,890 meters (Jamieson et al.,  
3926 2019). Additionally, the deepest ever sub diving recorded to date has registered the  
3927 presence of plastics on the ocean floor, at a depth of 10,927 meters (Street, 2019). It  
3928 is alarming to find plastics, which are materials with a history of less than a century  
3929 of large-scale production, already being so widespread in nature; with the potential  
3930 for extensive negative impacts, many of which have yet to be fully realised. Hence,  
3931 better plastic waste management and processing solutions are urgently required.

3932           Currently, the majority of plastic waste recycling is based on mechanical  
3933 recycling (collection, sorting, washing and grinding) (Ragaert et al., 2017). However,  
3934 the presence of organic and inorganic impurities in post-consumer plastic waste  
3935 presents a huge challenge for mechanical recycling (Drzyzga and Prieto, 2019). On  
3936 the other hand, chemical recycling has been applied as an alternative for improved  
3937 plastic waste management processes, in which the plastic polymers can be  
3938 converted into raw materials that can be used for the synthesis of chemicals, fuels,  
3939 or virgin plastics (Drzyzga and Prieto, 2019). Strictly chemical methods, however,  
3940 require the use of toxic chemicals and high temperatures, and can also be quite  
3941 costly (Awaja and Pavel, 2005; Wei and Zimmermann, 2017b). Therefore, enzymatic  
3942 hydrolysis of synthetic polyesters has been proposed as a potentially more efficient  
3943 and environment friendly method for the recycling of plastic waste (Drzyzga and  
3944 Prieto, 2019; Wei and Zimmermann, 2017b, 2017b).

3945           In the past decade, a number of bacterial enzymes capable of degrading  
3946 synthetic polyesters, including the widely used polyethylene terephthalate (PET),  
3947 have been identified (Kawai et al., 2019; Wei and Zimmermann, 2017b). These  
3948 enzymes are commonly classified as members of the cutinase, lipase and esterase  
3949 classes of enzymes, and to date have mainly been identified in thermophilic  
3950 actinomycetes, particularly in the genus *Thermobifida* (Silva et al., 2011; Wei and  
3951 Zimmermann, 2017b). More recently, in 2016, Yoshida and co-workers isolated a  
3952 bacterium from a PET plastic bottle recycling plant in Sakai, Japan, that was  
3953 capable of degrading and assimilating PET as its major energy and carbon source –  
3954 namely *Ideonella sakaiensis* 201-F6 (Yoshida et al., 2016). The protein identified as  
3955 being responsible for the hydrolysis of PET (ISF6\_4831) was then defined as a  
3956 PETase (or PET hydrolase) enzyme (EC 3.1.1.101) (Yoshida et al., 2016). The PETase  
3957 from *Ideonella sakaiensis* 201-F6 has been shown to possess a relatively higher  
3958 enzymatic activity and substrate specificity for PET than other previously



described PET hydrolases, in addition to the ability to degrade PET at moderate temperatures (around 30°C) (Joo et al., 2018; Yoshida et al., 2016). Since then, a number of studies have been undertaken in a concerted effort to characterise this enzyme and the underlying metabolic and biochemical processes involved in the degradation of PET (Chen et al., 2018; Han et al., 2017; Joo et al., 2018; Liu et al., 2018).

The *Streptomyces* genus, member of the Actinomycetales order, is well-known to produce compounds and enzymes of industrial and clinical interest, particularly antibiotics, for which it is considered the largest producer in the microbial world (Hwang et al., 2014; Ser et al., 2017; Spasic et al., 2018; Watve et al., 2001). Recent efforts to exploit the biotechnological potential of *Streptomyces* species have largely focused on the identification of bioactive small molecules and secondary metabolites biosynthetic gene clusters (Chevrette et al., 2019; Manteca and Yagüe, 2019). In this respect, the focus has started to shift towards *Streptomyces* isolates derived from varied niche environments, such as those isolated from the marine environment, which are still not well characterised and majorly unexplored organisms, when compared to the previously more commonly studied soil-derived isolates (Dharmaraj, 2010; Hassan et al., 2017; Jin et al., 2018; Xu et al., 2018).

*Streptomyces* isolates from soil ecosystems have also been studied for their synthetic polyesters-degrading capabilities (Calabia and Tokiwa, 2004; Shivilata and Satyanarayana, 2015). These include *Streptomyces* sp. strain MG (Tokiwa and Calabia, 2004) and *Streptomyces thermoviolaceus* (Chua et al., 2013), which can degrade polycaprolactone (PCL); together with *Streptomyces bangladeshensis* 77T-4, which degrades poly(D-3-hydroxybutyrate) (PHB) (Hsu et al., 2012). Given that marine *Streptomyces* sp. SNG9 had previously been reported to degrade PHB (Mabrouk and Sabry, 2001), coupled with the fact that marine *Streptomyces* isolates

are likely to have been exposed to plastics and/or microplastics in marine ecosystems – in particular those isolates which are associated with marine sponges (phylum *Porifera*), which filter large quantities of seawater (up to 24000 L of water per day/Kg sponge) on a daily basis to obtain nutrients (Food and Agriculture Organization of the United Nations, 2017; Godefroy et al., 2019; Taylor et al., 2007) – we reasoned that marine sponge-derived *Streptomyces* species may possess enzymes with an ability to degrade synthetic polymers. In this study, we screened a number of *Streptomyces* species, including both terrestrial and marine-derived isolates, using an *in silico*-based analysis to interrogate their genomes for potential PETase homologs. A candidate PETase-like gene was identified in *Streptomyces* sp. SM14 and enzyme activity was confirmed following heterologous expression of this gene in *Escherichia coli*. This is the first report of a PETase-like enzyme being identified in a marine sponge derived *Streptomyces* spp. isolates, and we believe that this study provides further insights into our current knowledge of this important class of enzymes.

### 4001 3. Material and methods

#### 4002 3.1. Data sets

4003 The reference data set was comprised of 15 amino acid sequences of  
4004 enzymes with previously demonstrated synthetic polyesters-degrading capabilities  
4005 (Table 1) (Danso et al., 2018; Joo et al., 2018; Kawai et al., 2019; Wei and  
4006 Zimmermann, 2017b; Yoshida et al., 2016). A lipase from *Streptomyces exfoliatus*  
4007 (PDB ID: 1JFR), which is a cutinase-like enzyme (Kawai et al., 2019; Wei et al., 1998),  
4008 was also included in the reference data set (Table 1), which although not possessing  
4009 demonstrated polyester-degrading activity, served as an outgroup for the  
4010 subsequent *in silico* analyses.

4011 The *Streptomyces* genomes data set comprised of 52 *Streptomyces* genome  
4012 sequences obtained from GenBank (Benson et al., 2018), including 23 genomes  
4013 from terrestrial isolates, and 29 from marine isolates (Table S1). Open reading  
4014 frames (ORFs) and their respective translated amino acid sequences were obtained  
4015 using Prokka (Seemann, 2014).

Gene name	Source	UniProt accession	GenBank accession	References
ISF6_483 (IsPETase)	<i>Ideonella sakaiensis</i> strain 201-F6	A0A0K8P6T7	GAP38373	Yoshida et al., 2016
Cut190	<i>Saccharomonospora viridis</i>	W0TJ64	BAO42836	Kawai et al., 2014
Tcur_1278	<i>Thermomonospora curvata</i> DSM 43183	D1A9G5	ACY96861	Chertkov et., 2011 Wei et al., 2014
Tha_Cut1	<i>Thermobifida alba</i>	E9LVH7	ADV92525	Ribitsch et al., 2012a
Thh_Est	<i>Thermobifida halotolerans</i>	H6WX58	AFA45122	Ribitsch et al., 2012b
Thc_Cut1	<i>Thermobifida cellulosilytica</i>	E9LVH8	ADV92526	Herrero Acero et al., 2011
Thc_Cut2	<i>Thermobifida cellulosilytica</i>	E9LVH9	ADV92527	Herrero Acero et al., 2011
Thf42_Cut1	<i>Thermobifida fusca</i>	E9LVI0	ADV92528	Herrero Acero et al., 2011
cut-1.KW3	<i>Thermobifida fusca</i>	E5BBQ2	CBY05529	Herrero Acero et al., 2011
cut-2.KW3	<i>Thermobifida fusca</i>	E5BBQ3	CBY05530	Herrero Acero et al., 2011
LCC	Leaf-branch compost metagenome	G9BY57	AEV21261	Sulaiman et al., 2012
cut_1	<i>Thermobifida fusca</i>	G8GER6	AET05798	Hegde and Veeranki, 2013
cut_2	<i>Thermobifida fusca</i>	Q6A0I4	AET05799	Hegde and Veeranki, 2013
Tfu_0882	<i>Thermobifida fusca</i> XY	Q47RJ7	AAZ54920	Chen et al., 2010
Tfu_0883	<i>Thermobifida fusca</i> XY	Q47RJ6	AAZ54921	Chen et al., 2010
Lipase (1JFR)	<i>Streptomyces exfoliatus</i>	Q56008	AAB51445	Wei et al., 1998

4016 **Table 1:** Reference data set comprising of 15 PETase-like enzymes with  
4017 demonstrated PET-degrading activity, including the ISF6\_483 protein from  
4018 *Ideonella sakaiensis* strain 201-F6 (IsPETase), and additionally the cutinase-like  
4019 lipase from *Streptomyces exfoliatus* (PDB ID: 1JFR).

### 3.2. Bacterial strains

*Streptomyces* strain SM14 was isolated from the sponge *Haliclona simulans* (class *Demospongiae*, order *Haplosclerida*, family *Chalinidae*) which was sampled by SCUBA diving at a depth of 15 m in Kilkieran Bay, Galway, Ireland (N 53°18'56.6", W 09°40'08.4") as previously described (Kennedy et al., 2014). The NEB® 5-alpha and the BL21(DE3) competent *E. coli* cells were obtained from New England Biolabs Inc., USA.

### 3.3. Protein homology search and phylogeny analysis

Potential PETase-like proteins were identified in the *Streptomyces* genomes data set by performing an homology search using BLASTP (e-value threshold of 1e-30, maximum subject sequence length of 400 aa) (Altschul et al., 1990; Camacho et al., 2009). Protein alignments were performed using Muscle (Edgar, 2004), and phylogeny analysis was performed using MEGA X (maximum likelihood statistical method; 500 bootstrap replications; LG+G+F model) (Kumar et al., 2018).

### 3.4. PCL plate clearing assay

Polycaprolactone (PCL) plate clearing assays were performed based on previously described studies (Murphy et al., 1996; Nawaz et al., 2015; Nishida et al., 1998; Nishida and Tokiwa, 1993). PCL with an average molecular weight of 80,000 was used (Sigma-Aldrich®). PCL emulsion was prepared with 1% m/v of PCL in acetone, at 50°C with magnetic stirring. Water, agar (1.5% m/v) and LB medium (2% m/v) were added to the emulsion, at 50°C with magnetic stirring until the acetone evaporated. The medium was then autoclaved and poured into plates. Strains were inoculated onto the plates and incubated at 28°C for up to 12 days. For enzyme activity assessment using *E. coli* as the heterologous host, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the medium at a concentration of 0.5 mM and plates were incubated at 28°C for up to 4 days. As a negative control, *E.*

4046 *coli* BL21(DE3) containing the pET-20b(+) plasmid without the insert did not show  
 4047 any PCL-degrading activity (Figure S1). It has recently been reported that *E. coli*  
 4048 BL21(DE3) can be employed as a host system in screens for polyesterase activity, as  
 4049 it does not possess PCL-degrading capabilities (Molitor et al., 2019). Additionally,  
 4050 12 other marine sponge-derived *Streptomyces* isolates were also assayed for  
 4051 polyesterase activity using the PCL plate clearing assay. SM1, SM3, SM4, SM7, SM8,  
 4052 SM9, SM11, SM13, SM17, and FMC008 which had previously been isolated from  
 4053 the marine sponge *Haliclona simulans* (Kennedy et al., 2009), together with  
 4054 B188M101 and B226SN101 isolated from the deep sea sponges *Lissodendoryx*  
 4055 *diversichela* and *Inflatella pellicula* respectively (Jackson et al., 2018); were grown  
 4056 in LB medium + 1% PCL emulsion at 28°C for 12 days (data not shown).

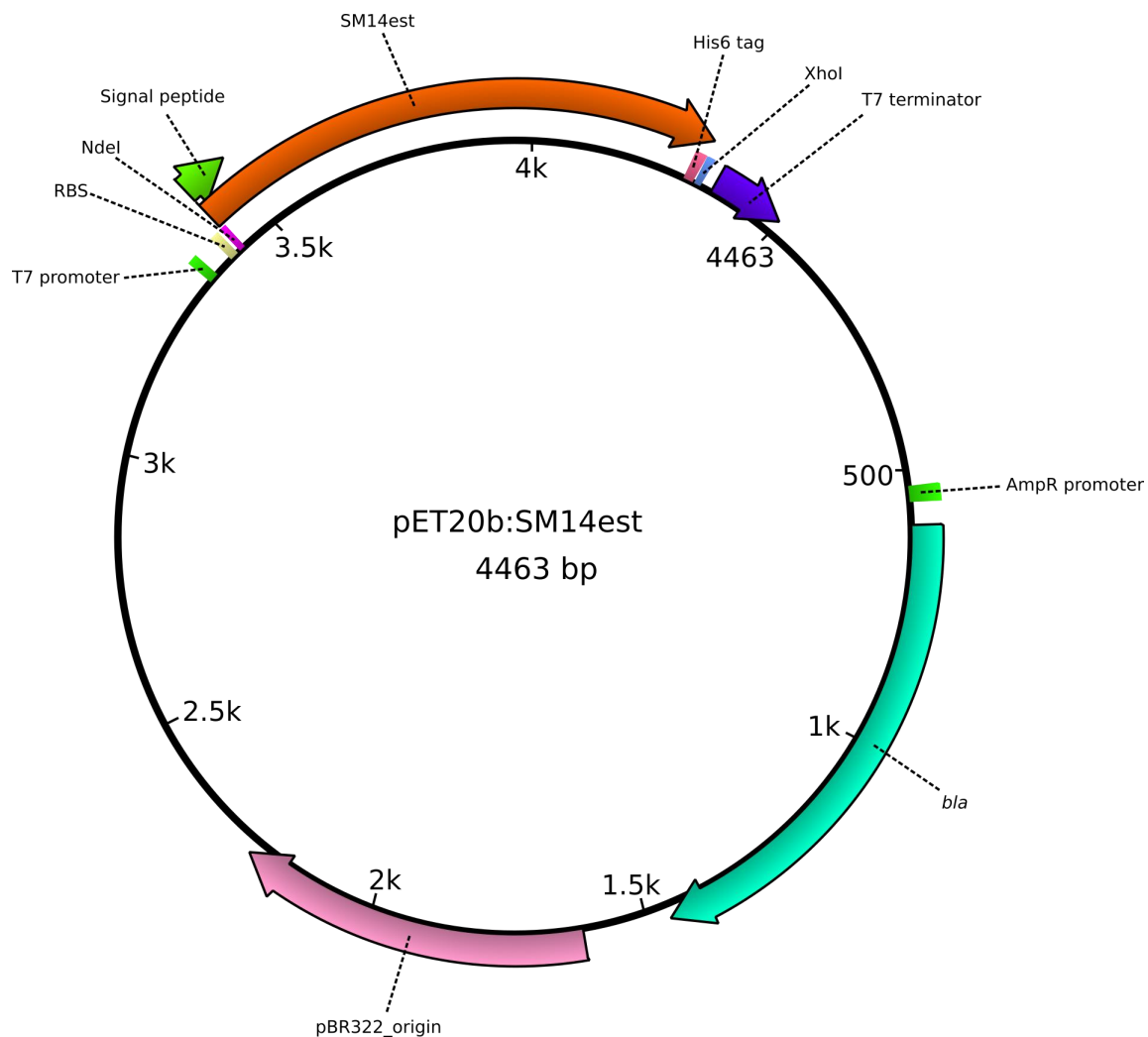
### 4057 3.5. Protein structure analysis, modelling and molecular docking

4058 Amino acid sequence analysis was performed and graphically represented  
 4059 using ESPript 3.0 (Robert and Gouet, 2014). *In silico* protein structure prediction  
 4060 was performed using the SWISS-DOCK webserver (Benkert et al., 2011; Bertoni et  
 4061 al., 2017; Bienert et al., 2017; Guex et al., 2009; Waterhouse et al., 2018), and the  
 4062 UCSF Chimera software was used for structure analysis and three-dimensional  
 4063 model rendering (Pettersen et al., 2004). Molecular docking experiments were  
 4064 performed using AutoDock Vina, MGLtools (<http://mgltools.scripps.edu/>),  
 4065 AutoDockTools (ADT) and UCSF Chimera (Pettersen et al., 2004; Sanner, 1999;  
 4066 Trott and Olson, 2009), with the model substrate BHET (Bis(2-hydroxyethyl)  
 4067 terephthalate, Zinc database ID: ZINC02040111) molecule as the ligand (Irwin et al.,  
 4068 2012; Irwin and Shoichet, 2005).

### 4069 3.6. Heterologous expression

4070 An *Escherichia coli* codon-optimised version of the PETase-like gene was  
 4071 designed, and was synthesised by Eurofins Genomics (Ebersberg, Germany). A 5'

4072 *NdeI* restriction site, a C-terminal His<sub>6</sub> tag, and a stop codon followed by a 3' *XhoI*  
4073 restriction site were added to the gene sequence. Alignment of the nucleotide  
4074 sequences of the original SM14est gene and the codon-optimised version is shown  
4075 in Figure S2. The synthetic gene was PCR amplified using Phusion Green High-  
4076 Fidelity DNA Polymerase (Thermo Scientific™) (primers and conditions detailed in  
4077 Table S2), and was subcloned into the pET-20b(+) plasmid (Novagen®), resulting in  
4078 the pET20b:SM14est vector construct (Figure 1), using the NEB® 5-alpha  
4079 competent *E. coli* (New England Biolabs Inc., USA) for vector construction and  
4080 maintenance. The signal peptide of the native protein was predicted using SignalP  
4081 5.0 (Almagro Armenteros et al., 2019), and it was maintained in the final construct.  
4082 The expression vector was then transformed into BL21(DE3) competent *E. coli*  
4083 (New England Biolabs Inc., USA) for heterologous protein expression.  
4084 Confirmation of the insert was performed via 1) restriction digestion of the plasmid  
4085 DNA with the *NdeI* and *XhoI* restriction enzymes followed by gel electrophoresis  
4086 analysis, and 2) via Sanger sequencing of the insert region of the plasmid,  
4087 amplified using the T7 standard vector primers (Table S2).



**Figure 1:** Graphical representation of the pET20b:SM14est plasmid, constructed for the heterologous expression of the SM14est protein in *E. coli*. The insert (SM14est) and other important features of the plasmid are represented and labelled accordingly.



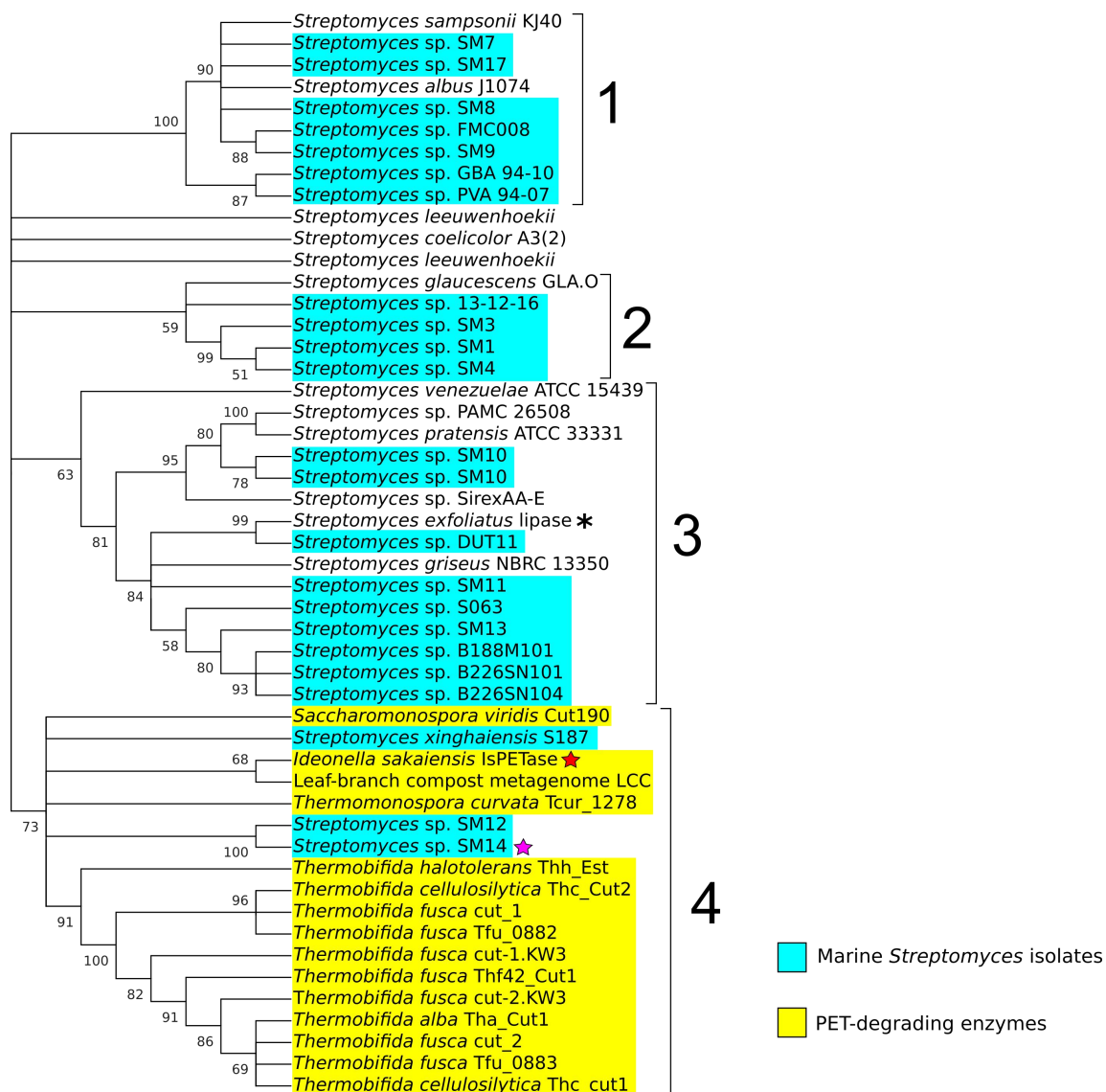
## 4. Results and Discussion

### 4.1. *In silico* screening of PETase-like proteins in *Streptomyces* genomes and enzyme activity assessment

Previous studies have identified enzymes with plastic-degrading capabilities which have been isolated from different organisms, amongst these were the *Ideonella sakaiensis* strain 201-F6, and isolates from the genera *Thermobifida*, *Thermomonospora*, and *Saccharomonospora* (Danso et al., 2018; Joo et al., 2018; Kawai et al., 2019; Wei and Zimmermann, 2017b). This class of enzymes is commonly referred to as PETase or PETase-like, due to their ability to hydrolyse PET, although at different levels of efficiency. From the bacterial PETase-like class of enzymes discovered to date, the PETase from *Ideonella sakaiensis* strain 201-F6 (referred to from now on as IsPETase) is the one that has received most attention from the scientific community; as it is considered the enzyme which displays the best PET hydrolase activity and substrate specificity which has been discovered thus far (Joo et al., 2018; Kawai et al., 2019; Yoshida et al., 2016). A total of 15 of the most prominent PETase-like enzymes were selected to build the reference data set (Table 1), which was then used to search for potential homologous proteins in the *Streptomyces* genomes data set. As described previously, the *Streptomyces* genomes data set comprised of 52 genome sequences, including 23 terrestrial *Streptomyces* isolates, and 29 isolates derived from the marine environment – 20 of which had previously been isolated by our group and for which genome sequence was available (Table S2) (Almeida et al., 2018; Jackson et al., 2018; Kennedy et al., 2009).

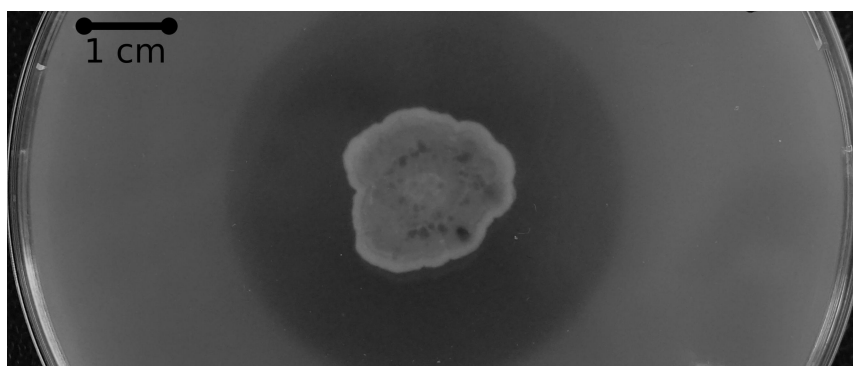
By applying a sequence similarity search approach using BLASTP (Altschul et al., 1990; Camacho et al., 2009), 34 homologous proteins from 32 *Streptomyces* strains were identified, of which the majority were from marine isolates (22 in total). These were then selected for further phylogeny analysis using MEGA X (Kumar et al., 2018). The amino acid sequences of the reference data set, and the *Streptomyces* potential PETase-like

4118 homologs were aligned using the Muscle program (Edgar, 2004), and a maximum  
 4119 likelihood phylogenetic tree was generated, with 500 bootstrap replicates. The resulting  
 4120 consensus phylogenetic tree, with a 50% bootstrap value cut-off, comprised of four main  
 4121 clades (Figure 2). Clade number 1 appeared to include *Streptomyces* isolates that were  
 4122 previously reported to share genetic similarity to the type strain *Streptomyces albus*  
 4123 J1074, all of which shared >99.50% 16S rRNA gene sequence similarity amongst each  
 4124 other (Ian et al., 2014; Zaburannyi et al., 2014). Clade number 2 showed less obvious  
 4125 similarities between all the members of the clade. While it included the strains SM1, SM3  
 4126 and SM4, that were isolated from the marine sponge *Haliclona simulans* and shared high  
 4127 similarity in their 16S rRNA sequences (>99%) (Jackson et al., 2018; Kennedy et al., 2009),  
 4128 it also included the sponge isolate *Streptomyces* sp. 13-12-16 that shared less 16s rRNA  
 4129 similarity with the aforementioned strains (~98%), and the soil isolate *Streptomyces*  
 4130 *glaucescens* strain GLA.O, which in addition to being isolated from a completely distinct  
 4131 environment, also shares <99% 16S rRNA similarity with the other strains in the clade.  
 4132 Clade number 3 appeared to be the most diverse, with *Streptomyces* isolated from varied  
 4133 sources, including: soil, marine sediment, and those isolated from marine sponges,  
 4134 lichens and insects (Bianchetti et al., 2013; Kennedy et al., 2009; Ohnishi et al., 2008; Shin  
 4135 et al., 2013; Xu et al., 2018). Clade 3 also included the lipase from *Streptomyces exfoliatus*  
 4136 (indicated with an asterisk in Figure 2), suggesting that these enzymes are likely to be  
 4137 cutinase-like lipases (Kawai et al., 2019; Wei et al., 1998). Most interesting, however, was  
 4138 clade number 4, which clearly included all of the PETase-like enzymes used in the  
 4139 reference data set, indicating that they may share similar evolutionary processes and  
 4140 history that differentiate them from the other proteins considered in this analysis, which  
 4141 may possibly have led to their ability to degrade synthetic polyesters. It is noteworthy  
 4142 that 3 proteins from marine *Streptomyces* isolates were also included in clade 4,  
 4143 specifically protein sequences from the *Streptomyces* sp. SM12, *Streptomyces* sp. SM14,  
 4144 and *Streptomyces xinghaiensis* S187 isolates, which is a strong indicator that these  
 4145 enzymes may possess plastic-degrading capabilities.



**Figure 2:** Phylogenetic tree of potential PETase homologs identified in the *Streptomyces* genomes, including terrestrial and marine (highlighted in cyan) *Streptomyces* isolates. The enzymes with known PET-degrading activity are highlighted in yellow. The red star indicates the *Ideonella sakaiensis* IsPETase, the pink star a PETase-like protein from the *Streptomyces* sp. SM14, and the asterisk the lipase from *Streptomyces exfoliatus*. The numbers in the branches indicate the percentage of bootstrap replicates (with a 50% cut-off from 500 replicates) in which the associated taxa clustered together.

4155 Subsequent amino acid sequence analysis showed that the SM12 and SM14  
4156 proteins are in fact identical, so additional analysis proceeded with the SM14 strain.  
4157 The enzyme activity was confirmed with a PCL plate clearing assay, in which the  
4158 SM14 strain was grown in LB medium + 1% PCL emulsion at 28°C for 12 days  
4159 (Figure 3). PCL is a synthetic polyester that has previously been used as a model  
4160 substrate to assess both PETase and cutinase enzymatic activities (Danson et al., 2018;  
4161 Nyssölä et al., 2013). The zone of clearing demonstrates the synthetic polyester-  
4162 degrading capability of the *Streptomyces* sp. SM14 isolate (Figure 3), which is  
4163 presumably due to the protein identified from the *in silico* screening (Figure 2).  
4164 Therefore, for the purposes of this study, the SM14 protein will from now on be  
4165 referred to as SM14est, as it is likely to be a potential polyesterase enzyme. The  
4166 SM14est gene sequence was deposited in the GenBank database under the  
4167 accession number BK010828.



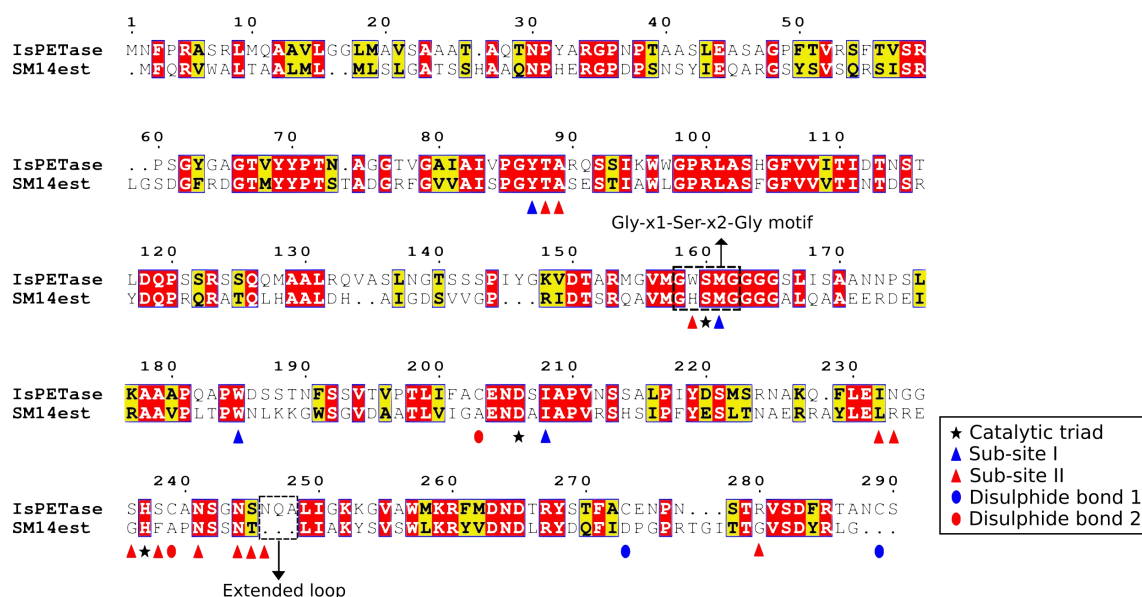
4168 **Figure 3:** PCL plate clearing assay with the *Streptomyces* sp. SM14  
4169 strain incubated at 28°C after 12 days.

## 4.2. Protein structure analysis

### 4.5.5. Amino acid sequence, conserved residues and domains

An amino acid sequence comparison between the SM14est and the IsPETase was performed, using previously described PETase enzyme sites as reference (Joo et al., 2018). The amino acid sequences were aligned using the Muscle algorithm in MEGA X (Edgar, 2004; Kumar et al., 2018), and the alignment and amino acid residues were analysed in ESPript 3.0 (Figure 4) (Robert and Gouet, 2014). The amino acid alignment showed that 41% of the amino acids in SM14est were identical to the IsPETase, and that an additional 19% of the SM14est amino acids shared similar biochemical properties to the IsPETase. The serine hydrolase motif Gly-x1-Ser-x2-Gly is conserved in both sequences (residues in IsPETase: Gly158-Trp159-Ser160-Met161-Gly162; residues in SM14est: Gly154-His155-Ser156-Met157-Gly158). The catalytic triad is also conserved in both sequences (residues in IsPETase: Ser160, Asp206, His237; and in SM14est: Ser156, Asp202, His234) (Figure 4 and Table S3), which is to be expected given that this catalytic triad has been shown to be crucial for enzymatic activity in this class of enzymes. In previous site-directed mutagenesis experiments performed with the IsPETase protein, substitution of any of the residues within the catalytic triad resulted in a complete disruption of the catalysis process (Joo et al., 2018; Liu et al., 2018). One major difference between the IsPETase and our SM14est is that the former possesses two disulphide bonds (the first between Cys273 and Cys289, and the second between Cys203 and Cys239), while the latter has none. Although disulphide bonds are generally related to higher protein thermostability, it has been proposed that the second disulphide bond of IsPETase is connected to its enzymatic activity, since it is positioned in close proximity to the enzyme's active sites, and substitution of this disulphide bond via site-directed mutagenesis experiments resulted in a drastic decrease in PET hydrolysis (Joo et al., 2018; Liu et al., 2018). However, the requirement of this

extra disulphide bond may be exclusive to the IsPETase, since other PETase-like cutinases have PET hydrolase activity and high thermostability without possessing this disulphide bond (Kawai et al., 2019).



**Figure 4:** Amino acid sequence alignment of the IsPETase and the SM14est proteins, generated using MEGA X, Muscle, and ESPrpt 3.0. Identical residues are highlighted in red boxes, and the ones with similar biochemical properties are highlighted in yellow boxes. The serine hydrolase motif Gly-x1-Ser-x2-Gly and the IsPETase extended loop are highlighted in black boxes; the catalytic triad is indicated by a star; the sub-site I and sub-site II are indicated by a blue and a red triangle, respectively. The disulphide bond 1 and disulphide bond 2 are indicated by blue and red ellipses, respectively.

Previous molecular docking analysis of the IsPETase, using a four-MHET molecule that would mimic PET (2-hydroxyethyl-(mono- hydroxyethyl terephthalate)<sub>4</sub>, or 2-HE(MHET)<sub>4</sub> as ligand; together with site-directed mutagenesis analysis, suggests that the enzyme possess two sub-sites (Joo et al., 2018). Sub-site I has been proposed to be responsible for the binding of the first

4214 MHET moiety, and thus for stabilisation of the ligand. Meanwhile, sub-site II has  
 4215 been proposed to be responsible for accommodating the other three MHET  
 4216 moieties, partially leading to the superior PET degradation by the IsPETase in  
 4217 comparison to other PETase-like enzymes. The sub-site I, which consists of 4  
 4218 residues, is conserved in both IsPETase and SM14est (residues in IsPETase: Tyr87,  
 4219 Met161, Trp185, Ile208; residues in SM14est: Tyr88, Met157, Trp181, Ile204) (Figure  
 4220 4 and Table S3). This implies that both enzymes have a similar mode of binding to  
 4221 the substrate. However, major differences exist between the two sequences in the  
 4222 12-residue sub-site II region (residues in IsPETase: Thr88, Ala89, Trp159, Ile232,  
 4223 Asn233, Ser236, Ser238, Asn241, Asn244, Ser245, Asn246, Arg280; residues in  
 4224 SM14est: Thr89, Ala90, His155, Leu229, Arg230, Gly233, Phe235, Asn238, Asn241,  
 4225 Thr242, GAP, Gly277) (Figure 4 and Table S3). These differences in the sub-site II  
 4226 region in both proteins could lead to different binding affinities to the moieties of  
 4227 the PET polymer. Another important difference between the two protein sequences  
 4228 is in the loop connecting  $\beta 8$  and  $\alpha 6$ , in which IsPETase appears to possess 3 extra  
 4229 amino acids in comparison to SM14est (Asn246, Gln247, Ala248) (Figure 4). This  
 4230 extended loop has been proposed to be an important structural feature of the  
 4231 IsPETase. When compared to a cutinase from *Thermobifida fusca* KW3, the  
 4232 extended loop seems to provide a conformation that allowed the formation of a  
 4233 continuous cleft on sub-site II; and hence accommodation of the third and fourth  
 4234 MHET moieties; therefore once more potentially explaining the superior enzymatic  
 4235 activity of IsPETase (Joo et al., 2018).

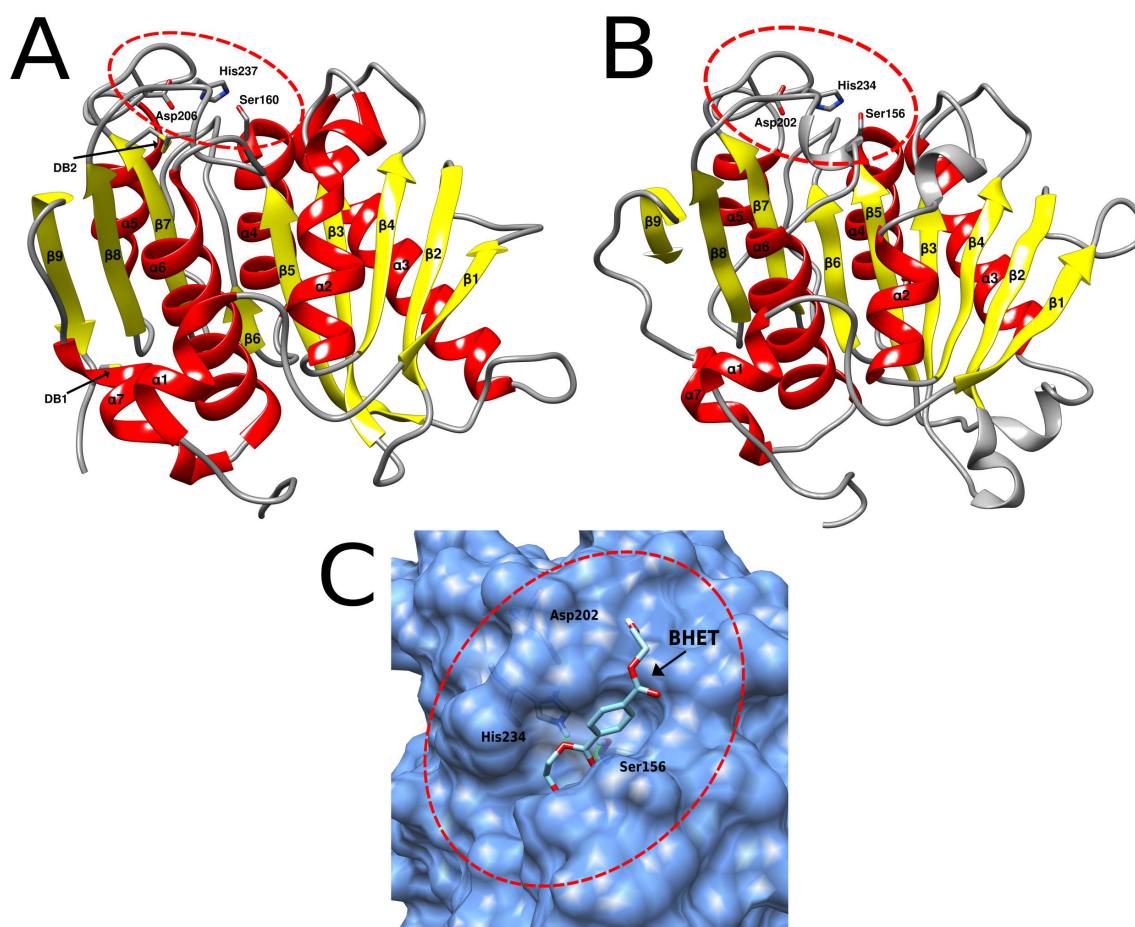
4236 It has been suggested that the aforementioned protein structural differences  
 4237 between the IsPETase and the cutinase from *Thermobifida fusca* KW3, namely the  
 4238 absence of two disulphide bonds; differences in the sub-site II residues and the lack  
 4239 of an extended loop; could result in a reduced efficiency in the degradation of PET  
 4240 when compared to the IsPETase, and it has been proposed that proteins with these

characteristics could be classified into the type I category of PETase-like enzymes, which also seems to be the case for the SM14est protein (Joo et al., 2018).

#### 4.5.6. Protein three-dimensional structure prediction and molecular docking

The function and stability of proteins is closely linked to its conformation, or folded/native state (Balcão and Vila, 2015; England and Haran, 2011; Lumry and Eyring, 1954). To provide further insights into the potential functionality and conformation of the SM14est protein, a three-dimensional structure of the protein was *in silico* predicted using SWISS-MODEL (Benkert et al., 2011; Bertoni et al., 2017; Bienert et al., 2017; Waterhouse et al., 2018). The cutinase 1 (Thc\_Cut1) from *Thermobifida cellulosilytica* (PDB ID: 5LUI) was used (Ribitsch et al., 2017) as a template for the model prediction, which generated a model with a GMQE score of 0.76 and a QMEAN Z-score of -1.76, indicating a reliable predicted model (Figure 5B). When compared to the structure of the IsPETase (Figure 5A), the predicted structure of the SM14est shows many similarities, with both belonging to the  $\alpha/\beta$  hydrolase superfamily (Hotelier et al., 2004; Nardini and Dijkstra, 1999; Ollis et al., 1992), displaying a similar arrangement of 9  $\beta$ -sheets and 7  $\alpha$ -helices (Figure 5). The arrangement of the catalytic triad residues is also quite similar, as highlighted in Figure 5A and Figure 5B, which may partially explain the synthetic polyester-degrading activity of these enzymes. With respect to observed potential differences between the two protein structures, the most striking differences; as previously shown in the amino acid sequence comparison, were the lack of disulphide bonds in the SM14est, and the absence of an extended loop between  $\beta$ 8 and  $\alpha$ 6. The latter, as previously mentioned, has been proposed to be linked to proper accommodation of the MHET moieties constituting the PET polymer, and therefore the superior enzymatic activity of the IsPETase (Joo et al., 2018).





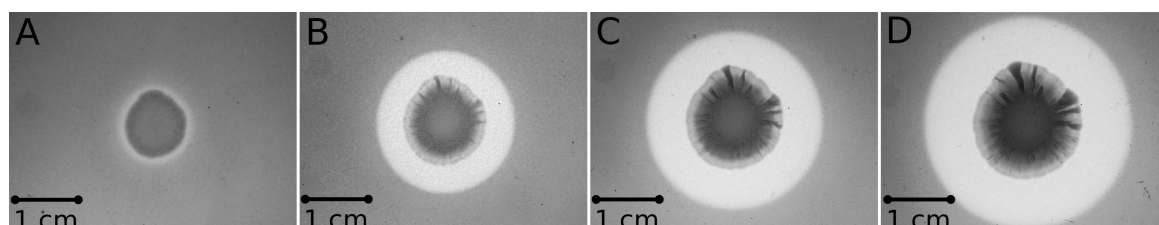
4266 **Figure 5:** Three-dimensional protein structure comparison and molecular docking.  
 4267 A) IsPETase three-dimensional structure (PDB ID: 5XJH), the catalytic triad (Ser160,  
 4268 Asp206, His237) is circled in red; the two disulphide bonds (DB1 and DB2) are  
 4269 indicated with arrows. B) Predicted three-dimensional structure of the SM14est  
 4270 protein, generated using SWISS-MODEL, with the crystal structure of cutinase 1  
 4271 from *Thermobifida cellulosilytica* as template (PDB ID: 5LUI). The catalytic triad  
 4272 (Ser156, Asp202, His234) is circled in red. C) Molecular docking simulation  
 4273 performed using AutoDock Vina, with BHET as substrate, indicated with an arrow,  
 4274 detailing the binding pocket, which is circled in red. The catalytic triad residues are  
 4275 shown as sticks. Structures were analysed and rendered using the UCSF Chimera  
 4276 software.

Molecular docking experiments were then performed, to analyse the likelihood of the SM14est enzyme possessing the capacity to bind plastics as substrates using an *in silico*-based approach; as well as determining the most probable binding mode of the protein to the ligand. To this end, the AutoDock Vina program was used for the protein-ligand molecular docking simulations (Trott and Olson, 2009), using the previously predicted SM14est structure and the BHET molecule as the ligand. BHET has previously been used as a model substrate for PET degradation both *in vitro* and *in silico* through molecular docking studies (Hantani et al., 2018; Joo et al., 2018; Liu et al., 2018). The molecular docking experiment generated 6 binding modes with energy  $\leq -5.0$  kcal/mol, similar to the binding modes energy values that had previously been described in similar molecular docking experiment performed with IsPETase and BHET (Liu et al., 2018). The binding modes were analysed in more detail under the UCSF Chimera software (Pettersen et al., 2004), and the best mode with the lowest binding energy is represented in Figure 5C, which highlights the binding pocket of the enzyme, the catalytic triad, and the proposed binding mode of the BHET molecule. The protein structure and molecular docking analyses results further emphasise the potential of the SM14est to degrade plastics, and highlight the structural features that may facilitate this enzymatic activity.

#### 4.3. Heterologous expression of SM14est

To determine whether the SM14est gene does possess polyesterase activity, the gene was firstly codon-optimized for *E. coli* to facilitate heterologous expression in this host. The gene was then cloned into the expression vector pET-20b(+), generating the construct pET20b:SM14est, in which the native signal peptide sequence was maintained (Figure 1). The construct was then transformed into *E. coli* BL21(DE3) and transformants were tested for activity by performing a PCL plate clearing assay (Figure 6). A halo of clearing, which is indicative of PCL

degradation, was observed following one day of incubation (Figure 6A), with the activity subsequently increasing after 2, 3 and 4 days (Figure 6B, C, and D, respectively). The negative control with *E. coli* BL21(DE3) containing the pET-20b(+) plasmid without the insert did not show any PCL-degrading activity (Figure S1).



**Figure 6:** PCL plate clearing assay with the *E. coli* BL21(DE3)(pET20b:SM14est) heterologous host, incubated at 28°C after A) 1 day; B) 2 days; C) 3 days; D) 4 days.

Another interesting observation was that the *E. coli* host successfully exported the heterologously expressed SM14est enzyme, when the native *Streptomyces* sp. SM14 signal peptide sequence was present in the expression construct (Figure 1, Figure 6). The signal peptide sequence was predicted to consist of the first 25 amino acids of the protein sequence, with a cleavage site probability of 0.9316, and to belong to the general secretory (Sec) pathway, with a likelihood value of 0.9608. In heterologous protein expression systems involving *E. coli*, successfully secretion and maintenance of the native protein confirmation can sometimes be challenging, and in this case potentially so, due to the fact the SM14est protein originates from such a distant host, *i.e.* a *Streptomyces* isolate (Freudl, 2018). Several *Streptomyces* genes encoding different enzymes have previously been heterologously expressed in *E. coli* (Spasic et al., 2018); including a xylanase from *S. mexicanus* HY-14 JQ943651 (Kim et al., 2014); a laccase from *S. coelicolor* (Sherif et al., 2013); a protease from *S. koyangensis* (Ben Elhoul et al., 2015); a glucose isomerase from *Streptomyces* sp. SK (Ben Hlima et al., 2013); and

4326 an esterase from *S. lividans* (Wang et al., 2016). In these cases, however, unlike with  
4327 the heterologous expression of the SM14est in *E. coli*, the native signal sequence  
4328 was not employed in the expression constructs.

4329         A number of different signal peptide sequences have previously been  
4330 employed to ensure the secretion of PETase and PETase-like enzymes from *E. coli*.  
4331 A PET carboxylesterase from *Thermobifida fusca* has been expressed and secreted  
4332 from *E. coli* using a pelB leader sequence (Oeser et al., 2010), while a PET hydrolase  
4333 has also previously been expressed and secreted from *Bacillus subtilis* using a  
4334 native PETase signal peptide (SP<sub>PETase</sub>) (Huang et al., 2018). In addition, a Sec-  
4335 dependent signal sequence from *E. coli* has also recently been used to express the  
4336 IsPETase, resulting in the production of the extracellular enzyme using *E. coli* (Seo  
4337 et al., 2019). The successful secretion of heterologously expressed PETases is  
4338 important not only for their subsequent purification, and biochemical  
4339 characterisation (Han et al., 2017; Joo et al., 2018; Liu et al., 2018), but also for their  
4340 potential development as recombinant PET-degrading microbes with utility in  
4341 environmental remediation. strategies. Therefore, it is interesting to note that we  
4342 were able to efficiently heterologous express the PETase-like SM14est enzyme from  
4343 *Streptomyces* sp. in *E. coli*, without the requirement to change its native signal  
4344 peptide sequence, with extracellular synthetic polyester-degrading activity being  
4345 observed in a PCL plate clearing assay. To our knowledge this is the first report of a  
4346 PETase-like enzyme being identified in a marine sponge-derived *Streptomyces* spp.  
4347 isolate, and we believe that the PETase-like SM14est enzyme will help provide  
4348 further insights into our current knowledge of this important class of synthetic  
4349 polyester-degrading enzymes.

4350 **5. Conclusion**

4351       Plastics such as the polyethylene terephthalate (PET) have been commonly  
4352 used in storage materials and in synthetic fabrics, and their resistance to  
4353 biodegradation has resulted in their accumulation in terrestrial and marine  
4354 ecosystems at an alarming rate. In an attempt to alleviate this problem, much  
4355 recent scientific interest has focused on the enzymatic hydrolysis of these types of  
4356 synthetic polyesters, including PET. While a number of PETase and PETase-like  
4357 enzymes have been identified and biochemically characterised, there is still much  
4358 to be learned about this class of enzymes. In addition, more information on their  
4359 structure, activity, and how widespread they are distributed in nature is required;  
4360 and if they can ultimately be improved using genetic and protein engineering and  
4361 applied in bioremediation strategies on an industrial scale.

4362       Although the *Streptomyces* genus is well studied with respect to the  
4363 production of bioactive compounds, less is known about their potential to produce  
4364 enzymes with synthetic polyester-degradation activities. In this study, based on an  
4365 *in silico* screening approach, we were able to identify a PETase-like enzyme,  
4366 namely SM14est, with synthetic polyester-degrading activity, which was isolated  
4367 from the marine sponge-derived strain *Streptomyces* sp. SM14, with enzyme  
4368 activity being confirmed *in vitro* with the heterologous expression of the protein in  
4369 *E. coli* using PCL plate clearing assays. Importantly, an active heterologously  
4370 expressed SM14est protein was secreted from *E. coli* with the native *Streptomyces*  
4371 SM14est signal peptide sequence. This will facilitate not only the future  
4372 biochemical characterisation of the protein, but also its potential utility in other  
4373 bioremediation-based applications targeting synthetic polyesters.

## 4374 6. References

- 4375 Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther,  
4376 O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions  
4377 using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi:10.1038/s41587-  
4378 019-0036-z.
- 4379 Almeida, E. L., Margassery, L. M., Kennedy, J., and Dobson, A. D. W. (2018). Draft  
4380 Genome Sequence of the Antimycin-Producing Bacterium *Streptomyces* sp.  
4381 Strain SM8, Isolated from the Marine Sponge *Haliclona simulans*. *Genome*  
4382 *Announc.* 6, e01535-17. doi:10.1128/genomeA.01535-17.
- 4383 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic  
4384 local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-  
4385 2836(05)80360-2.
- 4386 Awaja, F., and Pavel, D. (2005). Recycling of PET. *Eur. Polym. J.* 41, 1453–1477.  
4387 doi:10.1016/J.EURPOLYMJ.2005.02.005.
- 4388 Balcão, V. M., and Vila, M. M. D. C. (2015). Structural and functional stabilization of  
4389 protein entities: state-of-the-art. *Adv. Drug Deliv. Rev.* 93, 25–41.  
4390 doi:10.1016/j.addr.2014.10.005.
- 4391 Ben Elhoul, M., Zaraï Jaouadi, N., Rekik, H., Bejar, W., Boulkour Touiou, S., Hmidi,  
4392 M., et al. (2015). A novel detergent-stable solvent-tolerant serine thiol alkaline  
4393 protease from *Streptomyces koyangensis* TN650. *Int. J. Biol. Macromol.* 79,  
4394 871–882. doi:10.1016/j.ijbiomac.2015.06.006.
- 4395 Ben Hlima, H., Bejar, S., Riguet, J., Haser, R., and Aghajari, N. (2013). Identification  
4396 of critical residues for the activity and thermostability of *Streptomyces* sp. SK  
4397 glucose isomerase. *Appl. Microbiol. Biotechnol.* 97, 9715–9726.  
4398 doi:10.1007/s00253-013-4784-2.
- 4399 Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the  
4400 absolute quality of individual protein structure models. *Bioinformatics* 27,  
4401 343–350. doi:10.1093/bioinformatics/btq662.
- 4402 Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D.,  
4403 et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47.  
4404 doi:10.1093/nar/gkx1094.
- 4405 Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., and Schwede, T. (2017). Modeling  
4406 protein quaternary structure of homo- and hetero-oligomers beyond binary  
4407 interactions by homology. *Sci. Rep.* 7, 10480. doi:10.1038/s41598-017-09654-8.
- 4408 Bianchetti, C. M., Harmann, C. H., Takasuka, T. E., Hura, G. L., Dyer, K., and Fox, B.  
4409 G. (2013). Fusion of Dioxxygenase and Lignin-binding Domains in a Novel  
4410 Secreted Enzyme from Cellulolytic *Streptomyces* sp. SirexAA-E. *J. Biol. Chem.*

- 288, 18574–18587. doi:10.1074/jbc.M113.475848.
- Bienert, S., Waterhouse, A., de Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., et al. (2017). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 45, D313–D319. doi:10.1093/nar/gkw1132.
- Calabia, B. P., and Tokiwa, Y. (2004). Microbial degradation of poly(d-3-hydroxybutyrate) by a new thermophilic *Streptomyces* isolate. *Biotechnol. Lett.* 26, 15–19. doi:10.1023/B:BILE.0000009453.81444.51.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421.
- Chen, C.-C., Han, X., Ko, T.-P., Liu, W., and Guo, R.-T. (2018). Structural studies reveal the molecular mechanism of PETase. *FEBS J.* doi:10.1111/febs.14612.
- Chen, S., Su, L., Billig, S., Zimmermann, W., Chen, J., and Wu, J. (2010). Biochemical characterization of the cutinases from *Thermobifida fusca*. *J. Mol. Catal. B Enzym.* 63, 121–127. doi:10.1016/J.MOLCATB.2010.01.001.
- Chertkov, O., Sikorski, J., Nolan, M., Lapidus, A., Lucas, S., Del Rio, T. G., et al. (2011). Complete genome sequence of *Thermomonospora curvata* type strain (B9). *Stand. Genomic Sci.* 4, 13–22. doi:10.4056/sigs.1453580.
- Chevrette, M. G., Carlson, C. M., Ortega, H. E., Thomas, C., Ananiev, G. E., Barns, K. J., et al. (2019). The antimicrobial potential of *Streptomyces* from insect microbiomes. *Nat. Commun.* 10, 516. doi:10.1038/s41467-019-08438-0.
- Chua, T.-K., Tseng, M., and Yang, M.-K. (2013). Degradation of Poly(ε-caprolactone) by thermophilic *Streptomyces thermoviolaceus* subsp. *thermoviolaceus* 76T-2. *AMB Express* 3, 8. doi:10.1186/2191-0855-3-8.
- Danso, D., Schmeisser, C., Chow, J., Zimmermann, W., Wei, R., Leggewie, C., et al. (2018). New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. *Appl. Environ. Microbiol.* 84, e02773-17. doi:10.1128/AEM.02773-17.
- Dharmaraj, S. (2010). Marine *Streptomyces* as a novel source of bioactive substances. *World J. Microbiol. Biotechnol.* 26, 2123–2139. doi:10.1007/s11274-010-0415-6.
- Drzyzga, O., and Prieto, A. (2019). Plastic waste management, a matter for the “community”. *Microb. Biotechnol.* 12, 66–68. doi:10.1111/1751-7915.13328.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and

- high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.
- England, J. L., and Haran, G. (2011). Role of Solvation Effects in Protein Denaturation: From Thermodynamics to Single Molecules and Back. *Annu. Rev. Phys. Chem.* 62, 257–277. doi:10.1146/annurev-physchem-032210-103531.
- Food and Agriculture Organization of the United Nations (2017). Sponges and their role in the marine environment. Available at: <http://www.fao.org/3/a-i7775e.pdf> [Accessed July 31, 2019].
- Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb. Cell Fact.* 17, 52. doi:10.1186/s12934-018-0901-3.
- Geyer, R., Jambeck, J. R., and Law, K. L. (2017). Production, use, and fate of all plastics ever made. *Sci. Adv.* 3, e1700782. doi:10.1126/sciadv.1700782.
- Godefroy, N., Le Goff, E., Martinand-Mari, C., Belkhir, K., Vacelet, J., and Baghdiguian, S. (2019). Sponge digestive system diversity and evolution: filter feeding to carnivory. *Cell Tissue Res.*, 1–11. doi:10.1007/s00441-019-03032-8.
- Guex, N., Peitsch, M. C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, S162–S173. doi:10.1002/elps.200900140.
- Han, X., Liu, W., Huang, J.-W., Ma, J., Zheng, Y., Ko, T.-P., et al. (2017). Structural insight into catalytic mechanism of PET hydrolase. *Nat. Commun.* 8, 2106. doi:10.1038/s41467-017-02255-z.
- Hantani, Y., Imamura, H., Yamamoto, T., Senga, A., Yamagami, Y., Kato, M., et al. (2018). Functional characterizations of polyethylene terephthalate-degrading cutinase-like enzyme Cut190 mutants using bis(2-hydroxyethyl) terephthalate as the model substrate. *AIMS Biophys.* 5, 290–302. doi:10.3934/biophy.2018.4.290.
- Hassan, S. S. ul, Anjum, K., Abbas, S. Q., Akhter, N., Shagufta, B. I., Shah, S. A. A., et al. (2017). Emerging biopharmaceuticals from marine actinobacteria. *Environ. Toxicol. Pharmacol.* 49, 34–47. doi:10.1016/j.etap.2016.11.015.
- Hegde, K., and Veeranki, V. D. (2013). Production Optimization and Characterization of Recombinant Cutinases from *Thermobifida fusca* sp. NRRL B-8184. *Appl. Biochem. Biotechnol.* 170, 654–675. doi:10.1007/s12010-013-0219-x.
- Herrero Acero, E., Ribitsch, D., Steinkellner, G., Gruber, K., Greimel, K., Eiteljoerg, I., et al. (2011). Enzymatic Surface Hydrolysis of PET: Effect of Structural



- 4480 Diversity on Kinetic Properties of Cutinases from *Thermobifida*.  
4481 *Macromolecules* 44, 4632–4640. doi:10.1021/ma200949p.
- 4482 Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P., and Chatonnet, A. (2004).  
4483 ESTHER, the database of the  $\alpha/\beta$ -hydrolase fold superfamily of proteins.  
4484 *Nucleic Acids Res.* 32, 145D – 147. doi:10.1093/nar/gkh141.
- 4485 Hsu, K. J., Tseng, M., Don, T. M., and Yang, M. K. (2012). Biodegradation of poly( $\beta$ -  
4486 hydroxybutyrate) by a novel isolate of *Streptomyces bangladeshensis* 77T-4.  
4487 *Bot. Stud.* 53, 307–313.
- 4488 Huang, X., Cao, L., Qin, Z., Li, S., Kong, W., and Liu, Y. (2018). Tat-Independent  
4489 Secretion of Polyethylene Terephthalate Hydrolase PETase in *Bacillus subtilis*  
4490 168 Mediated by Its Native Signal Peptide. *J. Agric. Food Chem.* 66, 13217–  
4491 13227. doi:10.1021/acs.jafc.8b05038.
- 4492 Hwang, K.-S., Kim, H. U., Charusanti, P., Palsson, B. Ø., and Lee, S. Y. (2014).  
4493 Systems biology and biotechnology of *Streptomyces* species for the production  
4494 of secondary metabolites. *Biotechnol. Adv.* 32, 255–268.  
4495 doi:10.1016/j.biotechadv.2013.10.008.
- 4496 Ian, E., Malko, D. B., Sekurova, O. N., Bredholt, H., Rückert, C., Borisova, M. E., et  
4497 al. (2014). Genomics of Sponge-Associated *Streptomyces* spp. Closely Related  
4498 to *Streptomyces albus* J1074: Insights into Marine Adaptation and Secondary  
4499 Metabolite Biosynthesis Potential. *PLoS One* 9, e96719.  
4500 doi:10.1371/journal.pone.0096719.
- 4501 Irwin, J. J., and Shoichet, B. K. (2005). ZINC - A free database of commercially  
4502 available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182.  
4503 doi:10.1021/ci049714+.
- 4504 Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012).  
4505 ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* 52,  
4506 1757–1768. doi:10.1021/ci3001277.
- 4507 Jackson, S., Crossman, L., Almeida, E., Margassery, L., Kennedy, J., and Dobson, A.  
4508 (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene  
4509 Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates.  
4510 *Mar. Drugs* 16, 67. doi:10.3390/md16020067.
- 4511 Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., et al.  
4512 (2015). Plastic waste inputs from land into the ocean. *Science.* 347, 768–771.  
4513 doi:10.1126/science.1260352.
- 4514 Jamieson, A. J., Brooks, L. S. R., Reid, W. D. K., Pierniey, S. B., Narayanaswamy, B.  
4515 E., and Linley, T. D. (2019). Microplastics and synthetic particles ingested by  
4516 deep-sea amphipods in six of the deepest marine ecosystems on Earth. *R. Soc.*  
4517 *Open Sci.* 6, 180667. doi:10.1098/rsos.180667.

- 4518 Jin, J., Yang, X., Liu, T., Xiao, H., Wang, G., Zhou, M., et al. (2018). Fluostatins M–Q  
4519 Featuring a 6-5-6-6 Ring Skeleton and High Oxidized A-Rings from Marine  
4520 *Streptomyces* sp. PKU-MA00045. *Mar. Drugs* 16, 87. doi:10.3390/md16030087.
- 4521 Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., et al. (2018). Structural  
4522 insight into molecular mechanism of poly(ethylene terephthalate) degradation.  
4523 *Nat. Commun.* 9, 382. doi:10.1038/s41467-018-02881-1.
- 4524 Kawai, F., Kawabata, T., and Oda, M. (2019). Current knowledge on enzymatic PET  
4525 degradation and its possible application to waste stream management and  
4526 other fields. *Appl. Microbiol. Biotechnol.*, 1–16. doi:10.1007/s00253-019-09717-y.
- 4527 Kawai, F., Oda, M., Tamashiro, T., Waku, T., Tanaka, N., Yamamoto, M., et al. (2014).  
4528 A novel Ca<sup>2+</sup>-activated, thermostabilized polyesterase capable of hydrolyzing  
4529 polyethylene terephthalate from *Saccharomonospora viridis* AHK190. *Appl.*  
4530 *Microbiol. Biotechnol.* 98, 10053–10064. doi:10.1007/s00253-014-5860-y.
- 4531 Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., et al. (2009).  
4532 Isolation and Analysis of Bacteria with Antimicrobial Activities from the  
4533 Marine Sponge *Haliclona simulans* Collected from Irish Waters. *Mar.*  
4534 *Biotechnol.* 11, 384–396. doi:10.1007/s10126-008-9154-1.
- 4535 Kennedy, J., Flemer, B., Jackson, S. A., Morrissey, J. P., O’Gara, F., and Dobson, A. D.  
4536 W. (2014). Evidence of a Putative Deep Sea Specific Microbiome in Marine  
4537 Sponges. *PLoS One* 9, e91092. doi:10.1371/journal.pone.0091092.
- 4538 Kim, D. Y., Shin, D.-H., Jung, S., Lee, J. S., Cho, H.-Y., Bae, K. S., et al. (2014).  
4539 Biocatalytic properties and substrate-binding ability of a modular GH10  $\beta$ -1,4-  
4540 xylanase from an insect-symbiotic bacterium, *Streptomyces mexicanus* HY-14.  
4541 *J. Microbiol.* 52, 863–870. doi:10.1007/s12275-014-4390-8.
- 4542 Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X:  
4543 Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol.*  
4544 *Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.
- 4545 Lebreton, L., and Andrady, A. (2019). Future scenarios of global plastic waste  
4546 generation and disposal. *Palgrave Commun.* 5, 6. doi:10.1057/s41599-018-0212-  
4547 7.
- 4548 Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., et al. (2018). Protein Crystallography  
4549 and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate)  
4550 Hydrolase PETase from *Ideonella sakaiensis*. *ChemBioChem* 19, 1471–1475.  
4551 doi:10.1002/cbic.201800097.
- 4552 Lumry, R., and Eyring, H. (1954). Conformation Changes of Proteins. *J. Phys. Chem.*  
4553 58, 110–120. doi:10.1021/j150512a005.

- 4554 Mabrouk, M. M., and Sabry, S. A. (2001). Degradation of poly (3-hydroxybutyrate)  
4555 and its copolymer poly (3-hydroxybutyrate-co-3-hydroxyvalerate) by a marine  
4556 *Streptomyces* sp. SNG9. *Microbiol. Res.* 156, 323–335. doi:10.1078/0944-5013-  
4557 00115.
- 4558 Manteca, Á., and Yagüe, P. (2019). “*Streptomyces* as a Source of Antimicrobials:  
4559 Novel Approaches to Activate Cryptic Secondary Metabolite Pathways,” in  
4560 *Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity*  
4561 *Methods* (IntechOpen). doi:10.5772/intechopen.81812.
- 4562 Molitor, R., Bollinger, A., Kubicki, S., Loeschcke, A., Jaeger, K.-E., and Thies, S.  
4563 (2019). Agar plate-based screening methods for the identification of polyester  
4564 hydrolysis by *Pseudomonas* species. *Microb. Biotech.* 0, 1–11. doi:  
4565 10.1111/1751-7915.13418.
- 4566 Murphy, C. A., Cameron, J. A., Huang, S. J., and Vinopal, R. T. (1996). *Fusarium*  
4567 polycaprolactone depolymerase is cutinase. *Appl. Environ. Microbiol.* 62, 456–  
4568 60.
- 4569 Nardini, M., and Dijkstra, B. W. (1999).  $\alpha/\beta$  Hydrolase fold enzymes: the family  
4570 keeps growing. *Curr. Opin. Struct. Biol.* 9, 732–737. doi:10.1016/S0959-  
4571 440X(99)00037-8.
- 4572 Nawaz, A., Hasan, F., and Shah, A. A. (2015). Degradation of poly( $\epsilon$ -caprolactone)  
4573 (PCL) by a newly isolated *Brevundimonas* sp. strain MRL-AN1 from soil.  
4574 *FEMS Microbiol. Lett.* 362, 1–7. doi:10.1093/femsle/fnu004.
- 4575 Nishida, H., Suzuki, S., and Tokiwa, Y. (1998). Distribution of Poly( $\beta$ -propiolactone)  
4576 Aerobic Degrading Microorganisms in Different Environments. *J. Polym.*  
4577 *Environ.* 6, 43–58. doi:10.1023/A:1022878512743.
- 4578 Nishida, H., and Tokiwa, Y. (1993). Distribution of poly( $\beta$  -hydroxybutyrate) and  
4579 poly( $\epsilon$  -caprolactone) aerobic degrading microorganisms in different  
4580 environments. *J. Environ. Polym. Degrad.* 1, 227–233. doi:10.1007/BF01458031.
- 4581 Nyssölä, A., Pihlajaniemi, V., Järvinen, R., Mikander, S., Kontkanen, H., Kruus, K.,  
4582 et al. (2013). Screening of microbes for novel acidic cutinases and cloning and  
4583 expression of an acidic cutinase from *Aspergillus niger* CBS 513.88. *Enzyme*  
4584 *Microb. Technol.* 52, 272–278. doi:10.1016/j.enzmictec.2013.01.005.
- 4585 Oeser, T., Wei, R., Baumgarten, T., Billig, S., Föllner, C., and Zimmermann, W.  
4586 (2010). High level expression of a hydrophobic poly(ethylene terephthalate)-  
4587 hydrolyzing carboxylesterase from *Thermobifida fusca* KW3 in *Escherichia*  
4588 *coli* BL21(DE3). *J. Biotechnol.* 146, 100–104. doi:10.1016/j.jbiotec.2010.02.006.
- 4589 Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., et al. (2008).  
4590 Genome sequence of the streptomycin-producing microorganism  
4591 *Streptomyces griseus* IFO 13350. *J. Bacteriol.* 190, 4050–60.

- doi:10.1128/JB.00204-08.
- Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., et al. (1992). The  $\alpha/\beta$  hydrolase fold. *Protein Eng. Des. Sel.* 5, 197–211. doi:10.1093/protein/5.3.197.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/jcc.20084.
- PlasticsEurope (2018). Plastics – the Facts 2018. *PlasticsEurope*. Available at: <https://www.plasticseurope.org/en/resources/publications/619-plastics-facts-2018> [Accessed May 14, 2019].
- Ragaert, K., Delva, L., and Van Geem, K. (2017). Mechanical and chemical recycling of solid plastic waste. *Waste Manag.* 69, 24–58. doi:10.1016/J.WASMAN.2017.07.044.
- Ribitsch, D., Acero, E. H., Greimel, K., Eiteljoerg, I., Trotscha, E., Freddi, G., et al. (2012a). Characterization of a new cutinase from *Thermobifida alba* for PET-surface hydrolysis. *Biocatal. Biotransformation* 30, 2–9. doi:10.3109/10242422.2012.644435.
- Ribitsch, D., Herrero Acero, E., Greimel, K., Dellacher, A., Zitzenbacher, S., Marold, A., et al. (2012b). A New Esterase from *Thermobifida halotolerans* Hydrolyses Polyethylene Terephthalate (PET) and Polylactic Acid (PLA). *Polymers (Basel)*. 4, 617–629. doi:10.3390/polym4010617.
- Ribitsch, D., Hromic, A., Zitzenbacher, S., Zartl, B., Gamerith, C., Pellis, A., et al. (2017). Small cause, large effect: Structural characterization of cutinases from *Thermobifida cellulosilytica*. *Biotechnol. Bioeng.* 114, 2481–2488. doi:10.1002/bit.26372.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi:10.1093/nar/gku316.
- Sanner, M. F. (1999). Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 17, 57–61.
- Santillo, D., Miller, K., and Johnston, P. (2017). Microplastics as contaminants in commercially important seafood species. *Integr. Environ. Assess. Manag.* 13, 516–521. doi:10.1002/ieam.1909.
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153.

- 4627 Seo, H., Kim, S., Son, H. F., Sagong, H.-Y., Joo, S., and Kim, K.-J. (2019). Production  
4628 of extracellular PETase from *Ideonella sakaiensis* using sec-dependent signal  
4629 peptides in *E. coli*. *Biochem. Biophys. Res. Commun.* 508, 250–255.  
4630 doi:10.1016/J.BBRC.2018.11.087.
- 4631 Ser, H.-L., Tan, L. T.-H., Law, J. W.-F., Chan, K.-G., Duangjai, A., Saokaew, S., et al.  
4632 (2017). Focused Review: Cytotoxic and Antioxidant Potentials of Mangrove-  
4633 Derived *Streptomyces*. *Front. Microbiol.* 8, 2065. doi:10.3389/fmicb.2017.02065.
- 4634 Sherif, M., Waung, D., Korbeci, B., Mavisakalyan, V., Flick, R., Brown, G., et al.  
4635 (2013). Biochemical studies of the multicopper oxidase (small laccase) from  
4636 *Streptomyces coelicolor* using bioactive phytochemicals and site-directed  
4637 mutagenesis. *Microb. Biotechnol.* 6, 588–597. doi:10.1111/1751-7915.12068.
- 4638 Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T.-J., Lee, J. E., et al. (2013).  
4639 Advantages of Single-Molecule Real-Time Sequencing in High-GC Content  
4640 Genomes. *PLoS One* 8, e68824. doi:10.1371/journal.pone.0068824.
- 4641 Shivlata, L., and Satyanarayana, T. (2015). Thermophilic and alkaliphilic  
4642 Actinobacteria: biology and potential applications. *Front. Microbiol.* 6, 1014.  
4643 doi:10.3389/fmicb.2015.01014.
- 4644 Silva, C., Da, S., Silva, N., Matama, T., Araujo, R., Martins, M., et al., (2011).  
4645 Engineering *Thermobifida fusca* cutinase with increased activity on polyester  
4646 substrates. *Biotechnol. J.* 6, 1230–1239. doi:10.1002/biot.201000391.
- 4647 Spasic, J., Mandic, M., Djokic, L., and Nikodinovic-Runic, J. (2018). *Streptomyces*  
4648 spp. in the biocatalysis toolbox. *Appl. Microbiol. Biotechnol.* 102, 3513–3536.  
4649 doi:10.1007/s00253-018-8884-x.
- 4650 Street, F. (2019). Deepest ever dive finds plastic bag at bottom of Mariana Trench.  
4651 *CNN*. Available at: [https://edition.cnn.com/travel/article/victor-vescovo-](https://edition.cnn.com/travel/article/victor-vescovo-deepest-dive-pacific/index.html)  
4652 [deepest-dive-pacific/index.html](https://edition.cnn.com/travel/article/victor-vescovo-deepest-dive-pacific/index.html) [Accessed May 14, 2019].
- 4653 Sulaiman, S., Yamato, S., Kanaya, E., Kim, J.-J., Koga, Y., Takano, K., et al. (2012).  
4654 Isolation of a Novel Cutinase Homolog with Polyethylene Terephthalate-  
4655 Degrading Activity from Leaf-Branch Compost by Using a Metagenomic  
4656 Approach. *Appl. Environ. Microbiol.* 78, 1556–1562. doi:10.1128/AEM.06725-11.
- 4657 Taylor, M. W., Radax, R., Steger, D., and Wagner, M. (2007). Sponge-associated  
4658 microorganisms: evolution, ecology, and biotechnological potential. *Microbiol.*  
4659 *Mol. Biol. Rev.* 71(2), 295–347. doi:10.1128/MMBR.00040-06.
- 4660 Tokiwa, Y., and Calabia, B. P. (2004). Review Degradation of microbial polyesters.  
4661 *Biotechnol. Lett.* 26, 1181–1189. doi:10.1023/B:BILE.0000036599.15302.e5.
- 4662 Trott, O., and Olson, A. J. (2009). AutoDock Vina: Improving the speed and

- 4663 accuracy of docking with a new scoring function, efficient optimization, and  
4664 multithreading. *J. Comput. Chem.* 31, NA-NA. doi:10.1002/jcc.21334.
- 4665 Wang, B., Wang, A., Cao, Z., and Zhu, G. (2016). Characterization of a novel highly  
4666 thermostable esterase from the Gram-positive soil bacterium *Streptomyces*  
4667 *lividans* TK64. *Biotechnol. Appl. Biochem.* 63, 334–343. doi:10.1002/bab.1465.
- 4668 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al.  
4669 (2018). SWISS-MODEL: homology modelling of protein structures and  
4670 complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427.
- 4671 Watve, M., Tickoo, R., Jog, M., and Bhole, B. (2001). How many antibiotics are  
4672 produced by the genus *Streptomyces*? *Arch. Microbiol.* 176, 386–390.  
4673 doi:10.1007/s002030100345.
- 4674 Wei, R., Oeser, T., Then, J., Kühn, N., Barth, M., Schmidt, J., et al. (2014). Functional  
4675 characterization and structural modeling of synthetic polyester-degrading  
4676 hydrolases from *Thermomonospora curvata*. *AMB Express* 4, 44.  
4677 doi:10.1186/s13568-014-0044-9.
- 4678 Wei, R., and Zimmermann, W. (2017b). Biocatalysis as a green route for recycling  
4679 the recalcitrant plastic polyethylene terephthalate. *Microb. Biotechnol.* 10,  
4680 1302–1307. doi:10.1111/1751-7915.12714.
- 4681 Wei, R., and Zimmermann, W. (2017a). Microbial enzymes for the recycling of  
4682 recalcitrant petroleum-based plastics: how far are we? *Microb. Biotechnol.* 10,  
4683 1308–1322. doi:10.1111/1751-7915.12710.
- 4684 Wei, Y., Swenson, L., Castro, C., Derewenda, U., Minor, W., Arai, H., et al. (1998).  
4685 Structure of a microbial homologue of mammalian platelet-activating factor  
4686 acetylhydrolases: *Streptomyces exfoliatus* lipase at 1.9 Å resolution. *Structure* 6,  
4687 511–519. doi:10.1016/S0969-2126(98)00052-5.
- 4688 Xu, X.-N., Chen, L.-Y., Chen, C., Tang, Y.-J., Bai, F.-W., Su, C., et al. (2018). Genome  
4689 Mining of the Marine Actinomycete *Streptomyces* sp. DUT11 and Discovery of  
4690 Tunicamycins as Anti-complement Agents. *Front. Microbiol.* 9, 1318.  
4691 doi:10.3389/fmicb.2018.01318.
- 4692 Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., et al.  
4693 (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate).  
4694 *Science*. 351, 1196–1199. doi:10.1126/science.aad6359.
- 4695 Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., and Luzhetskyy, A. (2014).  
4696 Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC*  
4697 *Genomics* 15, 97. doi:10.1186/1471-2164-15-97.

## 4698 7. Supplementary material

Organism name	GenBank accession number	Source
<i>Streptomyces alboflavus</i> strain MDJK44	CP021748.1	Non-marine
<i>Streptomyces albus</i> strain BK3-25	NZ_CP016825.1	Non-marine
<i>Streptomyces albus</i> strain DSM 41398	NZ_CP010519.1	Non-marine
<i>Streptomyces albus</i> J1074	NC_020990.1	Non-marine
<i>Streptomyces globisporus</i> C-1027	CP013738.1	Non-marine
<i>Streptomyces griseus</i> subsp. <i>griseus</i> NBRC 13350	NC_010572.1	Non-marine
<i>Streptomyces</i> sp. PAMC26508	NC_021055.1	Non-marine
<i>Streptomyces pratensis</i> ATCC 33331	CP002475.1	Non-marine
<i>Streptomyces sampsonii</i> strain KJ40	NZ_CP016824.1	Non-marine
<i>Streptomyces avermitilis</i> MA-4680	BA000030.4	Non-marine
<i>Streptomyces bingchengensis</i> BCW-1	CP002047.1	Non-marine
<i>Streptomyces cattleya</i> DSM 46488	CP003219.1	Non-marine
<i>Streptomyces coelicolor</i> A3(2)	AL645882.2	Non-marine
<i>Streptomyces collinus</i> Tu 365	CP006259.1	Non-marine
<i>Streptomyces glaucescens</i> strain GLA.O	CP009438.1	Non-marine
<i>Streptomyces</i> sp. SirexAA-E	NC_015953.1	Non-marine
<i>Streptomyces leeuwenhoekii</i>	LN831790.1	Non-marine
<i>Streptomyces lincolnensis</i> strain NRRL 2936	CP016438.1	Non-marine
<i>Streptomyces noursei</i> ATCC 11455	CP011533.1	Non-marine
<i>Streptomyces rimosus</i> subsp. <i>rimosus</i> strain NRRL ISP-5260	NYR01000001.1	Non-marine
<i>Streptomyces scabiei</i> 87.22	FN554889.1	Non-marine
<i>Streptomyces venezuelae</i> strain ATCC 15439	CP013129.1	Non-marine
<i>Streptomyces vietnamensis</i> strain GIM4.0001	CP010407.1	Non-marine
<i>Streptomyces</i> sp. 13-12-16	NCTE01000967.1	Marine
<i>Streptomyces</i> sp. CNQ-509	CP011492.1	Marine
<i>Streptomyces</i> sp. DUT11	CP025511.1	Marine
<i>Streptomyces</i> sp. GBA 94-10	CM002271.1	Marine
<i>Streptomyces</i> sp. PVA 94-07	CM002273.1	Marine
<i>Streptomyces</i> sp. RV15	KQ949075.1	Marine
<i>Streptomyces</i> sp. S063	CP021707.1	Marine

<i>Streptomyces</i> sp. SCSIO 03032	CP021121.1	Marine
<i>Streptomyces xinghaiensis</i> S187	CP023202.1	Marine
<i>Streptomyces</i> sp. B188M101	GCA_002910985.1	Marine
<i>Streptomyces</i> sp. B226SN101	GCA_002910935.1	Marine
<i>Streptomyces</i> sp. B226SN104	Not deposited	Marine
<i>Streptomyces</i> sp. FMC008	Not deposited	Marine
<i>Streptomyces</i> sp. SM1	GCA_002910825.1	Marine
<i>Streptomyces</i> sp. SM3	Not deposited	Marine
<i>Streptomyces</i> sp. SM4	Not deposited	Marine
<i>Streptomyces</i> sp. SM5	GCA_002910895.1	Marine
<i>Streptomyces</i> sp. SM6	Not deposited	Marine
<i>Streptomyces</i> sp. SM7	Not deposited	Marine
<i>Streptomyces</i> sp. SM8	GCA_000299175.2	Marine
<i>Streptomyces</i> sp. SM9	GCA_002910795.1	Marine
<i>Streptomyces</i> sp. SM10	GCA_002910915.1	Marine
<i>Streptomyces</i> sp. SM11	GCA_002910905.1	Marine
<i>Streptomyces</i> sp. SM12	GCA_002910855.1	Marine
<i>Streptomyces</i> sp. SM13	GCA_002910875.1	Marine
<i>Streptomyces</i> sp. SM14	GCA_002910755.1	Marine
<i>Streptomyces</i> sp. SM17	GCA_002910725.2	Marine
<i>Streptomyces</i> sp. SM18	GCA_002910775.2	Marine
<i>Streptomyces</i> sp. SM19	Not deposited	Marine

---

**Table S1:** *Streptomyces* genome sequences employed in this study, obtained from the GenBank database when indicated. The isolation source is discriminated as marine and non-marine environment.

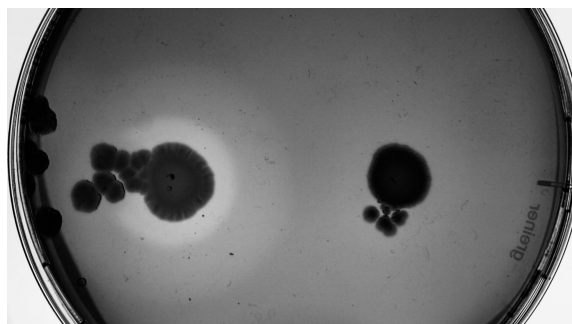


Primer name	Primer sequence	PCR conditions
SM14est_fw	5'-AAAAACATATGTTTCAGCGGGTCTGGGCGCTG-3'	Step 1: Incubate at 98°C for 30 s
		Step 2: Incubate at 98°C for 10 s
		Step 3: Incubate at 72°C for 30 s
SM14est_rev	5'-AAAAACTCGAGTTAGTGGTGATGGTGGTGATGGC-3'	Step 4: Cycle to step 2 for 35 times
		Step 5: Incubate at 72°C for 10 min
T7	5'-TAATACGACTCACTATAGG-3'	Step 1: Incubate at 98°C for 30 s
		Step 2: Incubate at 98°C for 10 s
		Step 3: Incubate at 53°C for 30 s
		Step 4: Incubate at 72°C for 15 s
T7_term	5'-CTAGTTATTGCTCAGCGGT-3'	Step 5: Cycle to step 2 for 35 times
		Step 6: Incubate at 72°C for 10 min

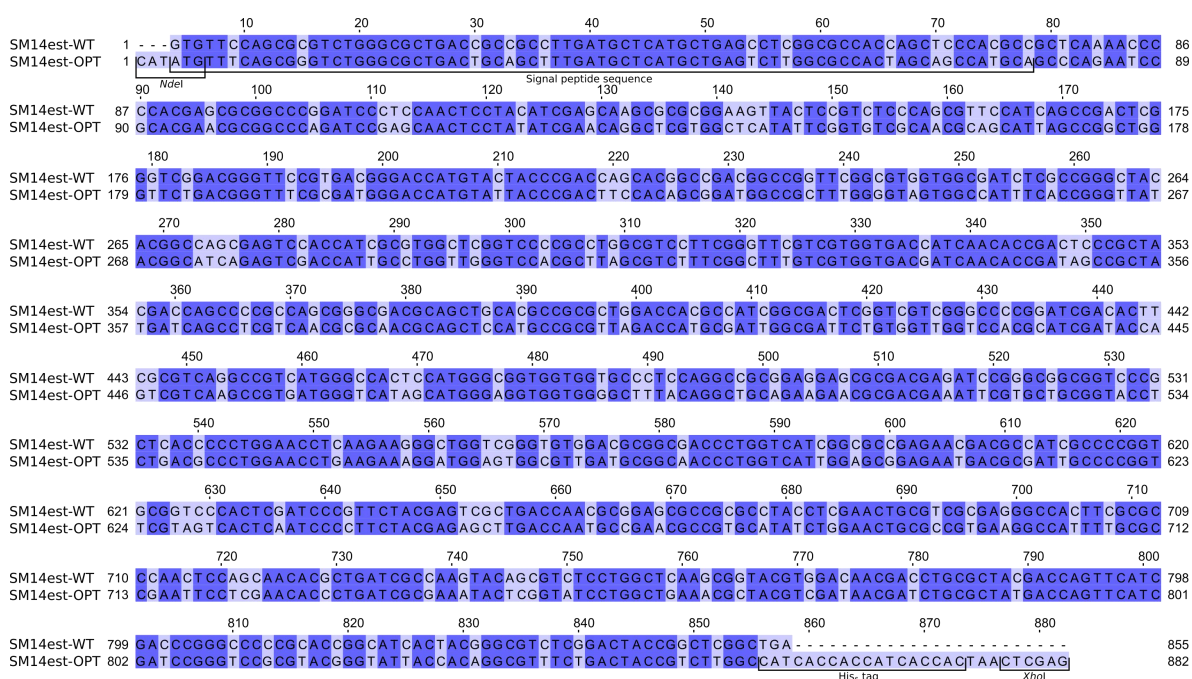
4702 **Table S2:** Primers and respective PCR amplification conditions  
 4703 employed in this study. The primer pair SM14est\_fw and  
 4704 SM14est\_rev was employed to subclone the synthetic SM14est gene  
 4705 sequence into the pET-20b(+) vector. The primer pair T7 and T7\_term  
 4706 (T7 standard primers) was employed to confirm the insert in the  
 4707 construct pET20b:SM14est.

	Residue in IsPETase	Residue in SM14est
<b>Catalytic triad</b>	Ser160	Ser156
	Asp206	Asp202
	His237	His234
<b>Sub-site I</b>	Tyr87	Tyr88
	Met161	Met157
	Trp185	Trp181
	Ile208	Ile204
<b>Sub-site II</b>	Thr88	Thr89
	Ala89	Ala90
	Trp159	His155*
	Ile232	Leu229**
	Asn233	Arg230*
	Ser236	Gly233*
	Ser238	Phe235*
	Asn241	Asn238
	Asn244	Asn241
	Ser245	Thr242**
	Asn246	_*
	Arg280	Gly277*

**Table S3:** Comparison of residues constituting the catalytic triad, sub-site I, and sub-site II in IsPETase and SM14est. Residue substitutions in SM14est are indicated with an asterisk, and those substitutions involving amino acids with similar biochemical properties are indicated with two asterisks.



**Figure S1:** PCL plate clearing assay at 28°C. Left: *E. coli* BL21(DE3)(pET20b:SM14est); right: *E. coli* BL21(DE3) negative control.



**Figure S2:** Nucleotide sequences alignment of the native SM14est protein codon usage (SM14est-WT) and the *E. coli* codon optimised synthetic version of the gene (SM14est-OPT). Darker boxes indicate identical nucleotides, while the lighter boxes indicate differences. Features such as the signal peptide sequence, and the engineered His<sub>6</sub> tag and restriction sites (*Nde*I and *Xho*I) are indicated accordingly.

# Chapter 5

## General Discussion and Concluding Remarks

## 4722 1. General discussion

### 4723 1.1. Marine sponge-derived *Streptomyces* species isolates: the importance of 4724 morphological, phenotypical, and genetic characterisation, and their potential 4725 as sources of novel bio-active compounds and enzymes

4726 Despite the fact that *Streptomyces* species isolates are well-known to have  
4727 the capacity to produce numerous bio-active molecules of pharmacological, clinical,  
4728 and biotechnological interest (Mohr, 2016; Watve et al., 2001), there is still a lack of  
4729 information regarding *Streptomyces* isolates derived from the marine environment.  
4730 Although recent research efforts have succeeded in identifying novel compounds  
4731 with relevant bio-activities from *Streptomyces* found associated with marine  
4732 sponges (Indraningrat et al., 2016), as shown in Chapter 1, very few studies have to  
4733 date focused on the characterisation of these isolates at a genetic, phenotypical and  
4734 morphological level (Ian et al., 2014; Tian et al., 2016).

4735 In Chapter 2, we analysed two *Streptomyces* spp. isolates, namely strains  
4736 SM17 and SM18, that were isolated from the marine sponge *Haliclona simulans*,  
4737 and which had previously been shown to possess bio-activity against clinically  
4738 relevant pathogens (Jackson et al., 2018; Kennedy et al., 2009). SM17 has been  
4739 shown to possess inhibitory activity against *Escherichia coli* NCIMB 12210, MRSA,  
4740 and *Candida* species; while SM18 inhibited the growth of MRSA and *Bacillus*  
4741 *subtilis* 1A40 (Jackson et al., 2018; Kennedy et al., 2009). In the study performed in  
4742 Chapter 2, we employed genomics and culture-based phenotypical characterisation  
4743 aimed at providing novel insights with respect to genetic traits that may be related  
4744 to environmental niche adaptations in SM17 and SM18, as well as their potential to  
4745 produce secondary metabolites (Almeida et al., 2019). State-of-the-art sequencing  
4746 technology was employed involving single molecule real-time sequencing (SMRT),  
4747 using the PacBio RSII sequencing platform; which allowed for the assembly of

4748 high-quality complete genome sequences for both SM17 and SM18 strains. This  
4749 resulted in, to our knowledge; two of the first complete genome sequences of  
4750 marine sponge-derived *Streptomyces* spp. isolates, which in itself represents an  
4751 important milestone and a data resource on these interesting bacteria, which will  
4752 be available to the wider scientific community. This also demonstrated that the  
4753 PacBio RSII sequencing platform provides excellent outcomes with respect to  
4754 sequencing the genomic DNA of uncharacterised environmental *Streptomyces*  
4755 isolates.

4756         A series of comparative analyses were performed in Chapter 2, between the  
4757 marine sponge-derived isolates SM17 and SM18 and their closest terrestrial  
4758 relatives, namely *Streptomyces albidoflavus* J1074 (referred to as *S. albus* J1074 in  
4759 Chapter 2) and *Streptomyces pratensis* ATCC 33331, respectively, which were  
4760 determined following phylogenomics-based analysis. Phenotypical differential  
4761 growth assessment showed that the marine sponge-derived isolates SM17 and  
4762 SM18 possessed an increased fitness to grow and differentiate in the presence of  
4763 salts in the growth medium, when compared to their terrestrial counterparts. This  
4764 was particularly true when comparing the SM17 and J1074 isolates, for which the  
4765 terrestrial J1074 appeared to struggle to grow in the medium enriched with salts,  
4766 while the marine SM17 thrived. Thus, it became clear that investigating genetic  
4767 differences between the marine and terrestrial isolates might provide insights with  
4768 respect to adaptations possessed by these marine-derived strains when compared  
4769 to their terrestrial counterparts, at a genetic level.

4770         Prediction of secondary metabolite biosynthetic gene clusters (BGCs) results  
4771 presented in Chapter 2 showed that the marine sponge-derived isolates SM17 and  
4772 SM18 possessed the potential to produce a range of diverse secondary metabolites,  
4773 some of which share similarities with known compounds, and that are also unique  
4774 in comparison to their terrestrial counterparts. This analysis further highlights the

4775 importance of exploring marine microbiomes, particularly those associated with  
4776 marine sponges, in the search for novel bioactive compounds.

4777         Interestingly, the genomic analysis in the search for BGCs showed that while  
4778 the SM18 isolate possessed siderophore-encoding BGCs, such as mirubactin and  
4779 coelichelin, it did not possess a BGC encoding for the production of the  
4780 desferrioxamines siderophores; which until our study were believed to be present  
4781 in all *Streptomyces* species (Cruz-Morales et al., 2017; Tierrafría et al., 2011). This  
4782 unprecedented characteristic could represent an adaptive and/or evolutionary trait,  
4783 which would be interesting to further investigate. For example, it would be  
4784 interesting to analyse the ability of the SM18 strain to grow in  
4785 presence/absence/different concentrations of iron in the culture medium, and how  
4786 it compares to other *Streptomyces* strains that possess desferrioxamines BGCs.

4787         An in-depth comparative genomics between SM17, SM18, J1074 and ATCC  
4788 33331 allowed us to identify a subset of genes that are commonly present in the  
4789 marine-derived isolates, and which are absent in their terrestrial counterparts.  
4790 These genes were *in silico* functionally characterised; and some were proposed to  
4791 represent environmental niche adaptations (ENA) that these marine isolates may  
4792 possess, at a genetic level. These adaptations included resistance to osmotic stress;  
4793 transcriptional regulation; symbiotic interactions; antimicrobial compounds  
4794 production and resistance; ABC transporters; and horizontal gene transfer (HGT)  
4795 and other potential defence-related features.

4796         Further genomic analysis regarding the ENA gene pool, including  
4797 *Streptomyces* isolates from other environments that were determined to be  
4798 phylogenetically related to the SM17 and SM18 strains; indicated the presence of  
4799 some of potential adaptation genes in other isolates. However, the majority of these  
4800 isolates were derived from environments in which similar stresses, as those that the  
4801 SM17 and SM18 are subjected to, could be experienced; since all of them were

isolated either from rhizosphere soil (which is known to present osmotic stress, for example), aquatic saline environments, or from symbiotic communities. Thus, this suggests that it is more likely that these adaptations may be related to specific stresses or environmental conditions, instead of general “marine” adaptations, in contrast to what has been suggested by previous studies (Ian et al., 2014; Penn and Jensen, 2012; Sun et al., 2018). A good example of this is the *nuo* operon, which has been proposed to be a marine adaptation in *Salinispora*, *Kocuria*, and other marine *Streptomyces* species (Ian et al., 2014; Penn and Jensen, 2012; Sun et al., 2018); however, this operon was also shown in Chapter 2 to be present in *Streptomyces* derived from rhizosphere soil and other aquatic saline environments, which could instead point to a trait involving osmoadaptation. Thus, in future studies, it would be interesting to assess the capability of these other environmental *Streptomyces* isolates to grow in the presence/different concentrations of salts in the culture medium.

Among the genes previously identified to be commonly present in the marine sponge-derived isolates SM17 and SM18, while absent in their terrestrial counterparts J1074 and ATCC 33331, a total of 7 of these genes were also absent in the other isolates analysed in the study. These consisted of: two genes encoding hypothetical proteins; a gene encoding a darcynin family of bacterial proteins with unknown function; a gene encoding an AAA family ATPase protein; the *yihX* gene, which seems to encode a phosphatase; the previously mentioned *tauB/ssuB* gene, which appears to encode an ABC transporter; and the *aldH* gene, which seems to encode a NADP-dependent aldehyde dehydrogenase. Although most of these genes seem to encode proteins with unknown functions, the previously mentioned TauB/SsuB and the YihX proteins may be related to a broader ability of the SM17 and SM18 strains for nutrient acquisition when compared to the other strains, specifically with respect to nitrate and sulfonate, and to phosphorus, respectively,



which are important resources for marine-derived bacteria (Hoppe HG, 2003; Karimi et al., 2018; Sebastian and Ammerman, 2009). Finally, interestingly, two of these genes, namely *aldH* and *tauB/ssuB*, were determined to be homologs of genes belonging to the marine isolates *Vibrio harveyi* ATCC 33843 and *Methanocaldococcus jannaschii* DSM 2661; result that further indicates their potential as adaptive traits to the conditions these isolates may be exposed to in marine environments.

## 1.2. Genome mining and the OSMAC approach as powerful methods for the identification of natural products in *Streptomyces* isolates

Chapter 3 followed up on some of the results presented in Chapter 2, in which the surugamides BGC (referred to as *sur* BGC) was predicted to be present in the *Streptomyces* sp. SM17 strain. More detailed phylogenomics analysis of this isolate, together with its closest relatives was performed, namely *Streptomyces albidoflavus* strain J1074, *S. albidoflavus* strain SM254, *S. sampsonii* strain KJ40, and *Streptomyces* sp. FR-008. In addition to sharing high similarities with respect to marker genes, i.e. the 16S rRNA gene, as well as the *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* genes, these isolates were also shown to share a considerably high number of genes, for which 5,177 genes (around 68% of their pan-genome) were determined to be conserved in all the isolates (i.e. the core genome), and 1,055 genes (around 14% of their pan-genome) were determined to be present in at least two isolates (i.e. the accessory genome). These results prompted us to propose the presence of the *albidoflavus* phylogroup, which has also been suggested in other studies (Labeda et al., 2014, 2017), and that all these strains are likely to be members of the same species. Hence, it is likely that the SM17 and FR-008 isolates in fact belong to the *S. albidoflavus* species, and that *Streptomyces sampsonii* KJ40 might have potentially been misassigned, which is in fact a common recurring issue in the taxonomy assignment of *Streptomyces* isolates. Good examples of this issue are coincidentally

4856 two strains analysed in this study that belong to the *albidoflavus* phylogroup,  
4857 namely *S. albidoflavus* J1074 and *S. albidoflavus* SM254, which have previously  
4858 been assigned as *S. albus* isolates, but which have more recently been re-assigned  
4859 as *S. albidoflavus* isolates (Hamm et al., 2017; Labeda et al., 2017).

4860 Interestingly, all of the members of the *albidoflavus* phylogroup that were  
4861 investigated in Chapter 3 do possess the BGC encoding for the production of  
4862 surugamides (*sur* BGC) (Ninomiya et al., 2016). Further analyses employing  
4863 phylogenetics and gene synteny showed that this BGC is highly conserved  
4864 amongst these isolates, and are in particular phylogenetically more similar between  
4865 the strains derived from aquatic-saline environments, namely *Streptomyces* sp.  
4866 SM17, *S. albidoflavus* LHW3101, and *S. albidoflavus* SM254. These results  
4867 provided evidence that there is a strong evolutionary pressure towards  
4868 maintaining the gene synteny – and even the reading frames of the main  
4869 biosynthetic genes – of the *sur* BGC, and also that this BGC potentially may in fact  
4870 have originated in the marine environment. Furthermore, the downstream region  
4871 to the main biosynthetic genes have been shown to be conserved in all the isolates,  
4872 with exception to the KJ40 strain. Whereas in the other strains the downstream  
4873 region to the *sur* BGC encompassed 1) a predicted multi-drug resistance (MDR)  
4874 transporter; 2) a predicted TetR/AcrR transcriptional regulator; 3) a hypothetical  
4875 protein; and 4) another predicted MDR transporter; while the downstream region  
4876 to the KJ40 *sur* BGC comprised of 1) a group of four hypothetical proteins, which  
4877 may represent pseudogene versions of the first MDR transporter identified in the  
4878 other isolates; 2) a predicted rearrangement hotspot (RHS) repeat protein; 3) a  
4879 hypothetical protein; and 4) a MDR transporter. Thus, it would be interesting to  
4880 evaluate the ability of the KJ40 isolate to produce surugamides and determine  
4881 whether these downstream genes do in fact play a role in their production.

4882           These results prompted us to investigate differences in the expression level  
4883 of surugamides when comparing a marine isolate (SM17) with a terrestrial isolate  
4884 (J1074) when grown in different culture conditions, employing an “One Strain  
4885 Many Compounds” (OSMAC)-based approach (Bode et al., 2002; Pan et al., 2019;  
4886 Romano et al., 2018). These two strains were grown in different complex media  
4887 (namely SYP-NaCl, YD, SY, P1, P2, P3, CH-F2, Oatmeal, and Sporulation media),  
4888 and their metabolic profile with respect to surugamide A production was analysed,  
4889 employing Liquid Chromatography–Mass Spectrometry (UPLC-DAD-HRMS and  
4890 UPLC-HRMS), High Performance Liquid Chromatography (HPLC) and Nuclear  
4891 Magnetic Resonance (NMR) spectroscopy.

4892           Surugamide A production was identified in the extracts from TSB, SYP-NaCl  
4893 and YD media, and interestingly SM17 was shown to be able to produce more  
4894 surugamide A than J1074 in all the conditions tested. In fact, production of  
4895 surugamide A by SM17 reached yields of up to >13-fold higher in YD medium, and  
4896 around 3-fold higher when grown in the SYP-NaCl medium when compared to  
4897 J1074. Identifying bacterial isolates that possess a higher capacity to produce  
4898 compounds of pharmacological interest under standard laboratory conditions is  
4899 certainly an important step towards the development of new drugs and industrial  
4900 processes.

4901           In both SM17 and J1074, the production of surugamide A was higher when  
4902 they were grown in SYP-NaCl medium. Interesting inferences can be made based  
4903 on these results. Since the media analysed use different carbon sources, it is  
4904 possible that carbon catabolite repression may play a major influence in the  
4905 production of surugamides, as it has also been reported to play for other secondary  
4906 metabolites (Brückner and Titgemeyer, 2002; Deutscher, 2008; Kremling et al., 2015;  
4907 Magnus et al., 2017; Romero-Rodríguez et al., 2017; Stülke and Hillen, 1999). The  
4908 SYP-NaCl medium has starch, a complex polysaccharide, as its main carbon source,

4909 while TSB and YD have glucose and dextrans, respectively. Additionally, the  
4910 presence of salts in the SYP-NaCl medium might have an influence on the  
4911 expression of the *sur* BGC. It has been reported that, for example, in the marine-  
4912 obligate actinobacteria *Salinispora arenicola*, different concentrations of salts in the  
4913 culture medium can result in different metabolic profiles; with different types  
4914 and/or quantities of secondary metabolites produced by this isolate being  
4915 influenced by the salts in the medium (Bose et al., 2015). Hence, it would be  
4916 interesting to, in future studies; assess the influence of different concentrations of  
4917 salts in the medium in the production of secondary metabolites by the SM17, J1074  
4918 and other *Streptomyces* isolates, particularly with respect to their potential  
4919 influence on the production of surugamides.

4920 All of these conclusions were only possible thanks to an initial genome  
4921 mining-based assessment of the SM17 genome, and subsequent culture-dependent  
4922 analyses coupled with analytical chemistry; which provided novel insights on the  
4923 distribution of the *sur* BGC in *Streptomyces* species, how conserved they are  
4924 genetically, and that carbon catabolite repression could be involved in the  
4925 production of surugamides. Thus, the results discussed in Chapter 3 ultimately  
4926 show that these approaches can also be very useful for the further characterisation  
4927 of compounds which have previously been isolated, such as the surugamides, and  
4928 to provide novel insights into regulatory mechanisms for the production of  
4929 secondary metabolites that can be ultimately useful for the optimisation of their  
4930 production in pharmacological and industrial settings.

### 4931 1.3. The potential of marine-derived *Streptomyces* isolates to harbour enzymes 4932 with synthetic polyesters-degrading activities

4933 Chapter 4 also clearly demonstrates how genome mining-based approaches,  
4934 coupled with culture-dependent assays and state-of-the-art synthetic biology can

4935 be applied when investigating the biotechnological potential of *Streptomyces*  
4936 isolates, however in a slightly different case scenario than the previous chapters.

4937         Instead of searching for BGCs encoding for the production of secondary  
4938 metabolites, alternatively, we employed an initial *in silico* screening based on  
4939 genome mining in a search for enzymes of potential biotechnological interest.  
4940 Specifically, enzymes with potential polyesterase activity were targeted, based on  
4941 the following rationale. Firstly, as it has been mentioned in Chapter 1, the issue of  
4942 plastic pollution is a serious problem, particularly when it comes to marine  
4943 ecosystems (Geyer et al., 2017; Jambeck et al., 2015). In fact, micro-plastics are now  
4944 believed to be ubiquitous in the oceans, and are likely to be consumed by marine  
4945 animals, and even possibly by marine sponges (Geyer et al., 2017; Lebreton and  
4946 Andrady, 2019). Thus, it is reasonable to assume that *Streptomyces* isolates found  
4947 in association with marine sponges are likely to have been in contact with these  
4948 ubiquitous micro-plastics; particularly given that some sponge species can filter up  
4949 to 20,000 litres of seawater kg<sup>-1</sup> h<sup>-1</sup> (dry wt) (Food and Agriculture Organization of  
4950 the United Nations, 2017; Taylor et al., 2007). Therefore, these microorganisms  
4951 could have developed – or inherited, e.g. via horizontal gene transfer –  
4952 mechanisms to degrade and assimilate these micro-plastics, potentially even using  
4953 them as energy source.

4954         A set of 15 nucleotide/amino acid sequences of known enzymes with PET  
4955 hydrolase activity was used as a reference (Kawai et al., 2019; Wei and  
4956 Zimmermann, 2017); and a set of 52 *Streptomyces* genome sequences from  
4957 terrestrial (23 genomes) and aquatic environments (29 genomes) was screened for  
4958 potential homologs to these reference PET hydrolase enzymes (referred to as  
4959 PETase-like enzymes). Phylogeny analysis of such potential PETase homologs  
4960 indicated that three protein sequences belonging to three marine-derived  
4961 *Streptomyces* spp. isolates, namely *Streptomyces* sp. SM14, *Streptomyces* sp. SM12,

4962 and *Streptomyces xinghaiensis* S187, shared high sequence similarities together  
4963 with close phylogenetic relatedness to the reference PETases. Due to two main  
4964 reasons, further analyses were performed with the SM14 strain and its potential  
4965 polyesterase enzyme. Firstly, further sequence analysis showed that the protein  
4966 sequences derived from the SM12 and SM14 strains were in fact identical, which  
4967 may not be surprising since these two isolates were determined to be very closely  
4968 related. Secondly, the SM14 strain had originally been isolated by our group and  
4969 genome sequence for the strain was available from previous works (Jackson et al.,  
4970 2018; Kennedy et al., 2009), hence this strain was readily available in our culture  
4971 collection, in contrast with the *S. xinghaiensis* S187, to which we had no ready  
4972 access to.

4973         The polyesterase activity of the *Streptomyces* sp. SM14 isolate was assessed  
4974 via a polycaprolactone (PCL) plate-clearing assay (Murphy et al., 1996; Nawaz et al.,  
4975 2015; Nishida et al., 1998; Nishida and Tokiwa, 1993), and confirmed when a  
4976 pronounced halo of clearing was observed. This prompted us to further investigate  
4977 the enzyme potentially responsible for the observed activity (namely SM14est)  
4978 employing both *in silico* and *in vitro* approaches.

4979         In Chapter 4, by employing *in silico* methods, the amino acid sequence and  
4980 the predicted three-dimensional structure of the SM14est protein was determined,  
4981 employing information from the currently best studied PETase, namely the  
4982 *Ideonella sakaiensis* 201-F6 PETase (referred to as IsPETase) (Joo et al., 2018;  
4983 Yoshida et al., 2016). This enabled us to provide some interesting insights with  
4984 respect to the structural characteristics of the SM14est protein and how they could  
4985 be related with its polyesterase activity, together with how differences that we  
4986 observed between SM14est and the IsPETase could be related to differences in the  
4987 enzyme's activities and affinities to plastics substrates. These inferences were also  
4988 based on previous studies that had experimentally determined the crystal structure

4989 of the IsPETase (Joo et al., 2018; Liu et al., 2018), and thus had identified some of  
4990 the structural features possessed by this enzyme that could, at least partially,  
4991 explain its superior activity and affinity to a PET substrate.

4992 In particular, it has been determined that the SM14est amino acid sequence  
4993 shared significant similarities to the IsPETase, with 41% amino acid identity in  
4994 addition to 19% of the amino acids sharing similar biochemical properties to those  
4995 of the IsPETase. The SM14est is like the IsPETase in that it displays a similar  
4996 arrangement of 9  $\beta$ -sheets and 7  $\alpha$ -helices, and thus may belong to the same  $\alpha/\beta$   
4997 hydrolase superfamily. The serine hydrolase motif (Gly-x1-Ser-x2-Gly) present in  
4998 the IsPETase was also determined to be conserved in the SM14est, i.e. Gly154-  
4999 His155-Ser156-Met157-Gly158, as well as the catalytic triad (Ser156, Asp202,  
5000 His234). The sub-site I of the IsPETase, which has been reported to be involved in  
5001 the stability of the substrate (Joo et al., 2018), also appears to be conserved in the  
5002 SM14est, encompassing the four residues Tyr88, Met157, Trp181, Ile204 in the  
5003 SM14est amino acid sequence.

5004 However, key sequence and structural differences were also identified when  
5005 comparing SM14est with the IsPETase. Major differences were observed in the 12-  
5006 residue sub-site II region, which has been suggested to provide a superior  
5007 accommodation of the moieties of the PET substrate (Joo et al., 2018). These  
5008 residues, in IsPETase, are Thr88, Ala89, Trp159, Ile232, Asn233, Ser236, Ser238,  
5009 Asn241, Asn244, Ser245, Asn246, Arg280; whereas in SM14est they have been  
5010 determined to encompass Thr89, Ala90, His155, Leu229, Arg230, Gly233, Phe235,  
5011 Asn238, Asn241, Thr242, an amino acid alignment gap, and Gly277. Also, the  
5012 IsPETase possesses an extended loop connecting  $\beta$ 8 and  $\alpha$ 6, which seems to be  
5013 absent in SM14est. This extended loop has been proposed to provide a higher  
5014 optimisation of the enzyme structure for an improved accommodation of the  
5015 substrate, and hence a superior enzymatic activity and affinity described in the

5016 IsPETase (Joo et al., 2018). Finally, two disulfide bonds that are present in the  
5017 IsPETase are absent in the SM14est, which may result in the SM14est being less  
5018 thermostable or in being less active, since one of these disulfide bonds (between  
5019 Cys203 and Cys239) is located in the vicinity of the catalytic site in the IsPETase  
5020 (Joo et al., 2018).

5021 Finally, in order to confirm that the previously observed polyesterase  
5022 activity by the SM14 strain was indeed being performed by the *in silico*-determined  
5023 SM14est enzyme we heterologously expressed a codon-optimised synthetic version  
5024 of the SM14est gene in *E. coli*. The activity was confirmed by employing a PCL  
5025 plate-clearing assay, in a similar fashion to the assay previously performed with the  
5026 wild type SM14 strain.

5027 Thus, in Chapter 4 an efficient method to identify PETase-like enzymes  
5028 based on genomic approaches is demonstrated, for which we were able to identify  
5029 and *in silico* characterise for the first time, to our knowledge, a polyesterase with  
5030 PCL-degrading activity derived from a marine sponge-associated *Streptomyces*  
5031 isolate. Secondly, an efficient method to heterologously express a PETase-like  
5032 enzyme from a *Streptomyces* strain is reported. The latter also represents a  
5033 potential improved activity of the enzyme, since the halos of clearing in the PCL  
5034 plate-clearing assays employing the heterologous host occurred in a significant  
5035 shorter period of time; with the activity observed in the heterologous *E. coli* host  
5036 system after 4 days being comparable to the activity observed after 12 days in the  
5037 native SM14 host. These activity levels would however need to be determined in  
5038 future studies to confirm this observation. Even more interestingly, however, is the  
5039 fact that the *E. coli* heterologous host was able to express and secrete a functional  
5040 polyesterase enzyme, even when the native *Streptomyces* signal peptide sequence  
5041 was employed in the synthetic gene construct used to express the *E. coli* codon-  
5042 optimised version of the SM14est gene. This novel finding was unexpected and to



our knowledge is likely to be unprecedented, as a native *Streptomyces* signal sequence being employed to export a functional *Streptomyces*-derived protein in an *E. coli* host system is highly unusual, mostly due to differences in the secretion pathways between the two genera. The impact of this finding could potentially go way beyond the scope of this study, and further investigations should be undertaken with this signal peptide, and how it could potentially be employed for the heterologous expression of other *Streptomyces*-derived enzymes in an *E. coli* host. Ideally, it would also have been interesting to analyse the effects of the SM14est enzyme in the degradation of PET and other recalcitrant plastics, but due to time and technical restrictions this was unfortunately not possible.

Nevertheless, the results presented in Chapter 4 represent new advances in our understanding of this extremely important PETase-like enzyme family. In particular, we report for the first time on an enzyme belonging to this family being present in a marine sponge-derived *Streptomyces* isolate. In addition by reporting on how genomic approaches coupled with simple *in vitro* assays and synthetic biology-based approaches can be used to identify PETase-like homologs; we open up the possibility of determining how potentially widespread in nature these enzymes might be; and hope that this may lead to the identification of other members of this enzyme family, with potential utility in various biotechnological applications and in particular in the degradation of synthetic polyesters.

## 2. Concluding remarks

Although recent efforts have shifted towards the exploration of marine-derived bacterial species aimed at the identification and isolation of novel compounds with pharmacological and biotechnological interest, marine sponge-associated *Streptomyces* species as a group are still largely underexplored. Throughout the work performed in this thesis, novel insights with respect to their genetics, biology and ecological niche lifestyle adaptations have been proposed, together with highlighting their outstanding potential as sources of potential bioactive compounds and enzymes of biotechnological and pharmacological interest.

Chapter 2 represents one of the very few studies published to date in regard to the characterisation of marine sponge-derived *Streptomyces* isolates, particularly from a genomics standpoint. Specifically, two marine sponge-derived isolates, namely *Streptomyces* sp. SM17 and *Streptomyces* sp. SM18, which had been subjected to state-of-the-art genome sequencing, represent two of the first complete genomes of marine sponge-derived *Streptomyces* available to date. These isolates were compared with their closest soil-derived counterparts, with a particular aim of identifying potential environmental niche adaptations possessed by these marine isolates. Additionally, the work performed here was crucial to proceeding with the analyses performed in the studies presented in the following chapters, with respect to assessing and setting the best approaches regarding genomics and bioinformatics-based pipelines for the analysis of our marine *Streptomyces* isolates.

In Chapter 3, using some of the genomics approaches established in Chapter 2, a new strain capable of producing surugamides was identified, namely *Streptomyces* sp. SM17. The surugamide family of compounds appears to be very promising due to their previously determined antitumor and antifungal properties,

5090 however very few studies focusing on this family of compounds have been  
5091 published to date (Kuranaga et al., 2018; Matsuda et al., 2019; Ninomiya et al., 2016;  
5092 Takada et al., 2013; Thankachan et al., 2019; Xu et al., 2017). Therefore, the results  
5093 presented in Chapter 3 are important in the context of gaining a better  
5094 understanding of this family of compounds, from both a genetic and evolutionary  
5095 standpoint, and to determine optimal conditions for their production. It also  
5096 reports on the identification of a *Streptomyces* strain (SM17) that appears to be able  
5097 of produce high levels of surugamide A in standard culture conditions, at least  
5098 when compared to the well-known strain *Streptomyces albidoflavus* J1074.

5099 Finally, in Chapter 4, the exploitation of the biotechnological potential of  
5100 marine-derived *Streptomyces* isolates employing genomics and bioinformatics-  
5101 based approaches is described. In particular, these approaches were employed for  
5102 the identification of novel enzymes potential with synthetic polyesters-degrading  
5103 activities. In this study, we were able to identify, for the first time, a PETase-like  
5104 polyesterase (SM14est) with PCL-degrading activity from a marine-sponge derived  
5105 *Streptomyces* isolate, namely *Streptomyces* sp. SM14. Further work with this  
5106 enzyme, such as an in-depth biochemical characterisation, together with further  
5107 assessment of its plastics-degrading capabilities using other substrates such as PET,  
5108 should be performed and would provide further insights into this family of  
5109 enzymes. Nevertheless, this study is fundamental in shedding new light on how  
5110 widespread in nature this family of enzymes might be, as well as reporting on an *in*  
5111 *silico* workflow that coupled with culture-dependent and state-of-the-art synthetic  
5112 biology methods can facilitate the identification of novel enzymes belonging to this  
5113 family. These enzymes may ultimately be used in bioremediation-based strategies  
5114 targeting plastic pollution, particularly in marine environments, or in novel  
5115 processes for plastic waste management.

5116           The work presented in this thesis ultimately helps to set new milestones  
5117 with respect to appropriate genomics-based approaches that can be employed to  
5118 exploit marine sponge-derived *Streptomyces* isolates, for biotechnological and  
5119 pharmacological applications, which have been demonstrated to be useful in this  
5120 work. In particular these approaches proved useful in the identification of a strain  
5121 which produces secondary metabolites of pharmacological interest, namely the  
5122 SM17 strain which produces surugamides, and in uncovering the polyester  
5123 degrading capabilities of marine *Streptomyces* isolates, with the identification, *in*  
5124 *silico* characterisation and heterologous expression of the SM14est polyesterase  
5125 derived from the SM14 strain.

5126 **3. References**

- 5127 Almeida, E. L. de, Carillo Rincon, F., Jackson, S. A., and Dobson, A. D. (2019).  
 5128 Comparative genomics of marine sponge-derived *Streptomyces* spp. isolates  
 5129 SM17 and SM18 with their closest terrestrial relatives provides novel insights  
 5130 into environmental niche adaptations and secondary metabolite biosynthesis  
 5131 potential. *Front. Microbiol.* 10, 1713. doi:10.3389/FMICB.2019.01713.
- 5132 Bode, H. B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big Effects from Small  
 5133 Changes: Possible Ways to Explore Nature's Chemical Diversity.  
 5134 *ChemBioChem* 3, 619. doi:10.1002/1439-7633(20020703)3:7<619::AID-  
 5135 CBIC619>3.0.CO;2-9.
- 5136 Bose, U., Hewavitharana, A., Ng, Y., Shaw, P., Fuerst, J., and Hodson, M. (2015). LC-  
 5137 MS-Based Metabolomics Study of Marine Bacterial Secondary Metabolite and  
 5138 Antibiotic Production in *Salinispora arenicola*. *Mar. Drugs* 13, 249–266.  
 5139 doi:10.3390/md13010249.
- 5140 Brückner, R., and Titgemeyer, F. (2002). Carbon catabolite repression in bacteria:  
 5141 Choice of the carbon source and autoregulatory limitation of sugar utilization.  
 5142 *FEMS Microbiol. Lett.* 209, 141–148. doi:10.1016/S0378-1097(02)00559-1.
- 5143 Cruz-Morales, P., Ramos-Aboites, H. E., Licona-Cassani, C., Selem-Mójica, N.,  
 5144 Mejía-Ponce, P. M., Souza-Saldívar, V., et al. (2017). Actinobacteria  
 5145 phylogenomics, selective isolation from an iron oligotrophic environment and  
 5146 siderophore functional characterization, unveil new desferrioxamine traits.  
 5147 *FEMS Microbiol. Ecol.* 93. doi:10.1093/femsec/fix086.
- 5148 Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria.  
 5149 *Curr. Opin. Microbiol.* 11, 87–93. doi:10.1016/j.mib.2008.02.007.
- 5150 Food and Agriculture Organization of the United Nations (2017). Sponges and their  
 5151 role in the marine environment. Available at: <http://www.fao.org/3/a-i7775e.pdf> [Accessed July 31, 2019].
- 5153 Geyer, R., Jambeck, J. R., and Law, K. L. (2017). Production, use, and fate of all  
 5154 plastics ever made. *Sci. Adv.* 3, e1700782. doi:10.1126/sciadv.1700782.
- 5155 Hamm, P. S., Caimi, N. A., Northup, D. E., Valdez, E. W., Buecher, D. C., Dunlap, C.  
 5156 A., et al. (2017). Western Bats as a Reservoir of Novel *Streptomyces* Species  
 5157 with Antifungal Activity. *Appl. Environ. Microbiol.* 83, e03057-16.  
 5158 doi:10.1128/AEM.03057-16.
- 5159 Hoppe HG (2003). Phosphatase activity in the sea. *Hydrobiologia* 493, 187–200.
- 5160 Ian, E., Malko, D. B., Sekurova, O. N., Bredholt, H., Rückert, C., Borisova, M. E., et  
 5161 al. (2014). Genomics of Sponge-Associated *Streptomyces* spp. Closely Related

- 5162 to *Streptomyces albus* J1074: Insights into Marine Adaptation and Secondary  
 5163 Metabolite Biosynthesis Potential. *PLoS One* 9, e96719.  
 5164 doi:10.1371/journal.pone.0096719.
- 5165 Indraningrat, A., Smidt, H., and Sipkema, D. (2016). Bioprospecting Sponge-  
 5166 Associated Microbes for Antimicrobial Compounds. *Mar. Drugs* 14, 87.  
 5167 doi:10.3390/md14050087.
- 5168 Jackson, S., Crossman, L., Almeida, E., Margassery, L., Kennedy, J., and Dobson, A.  
 5169 (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene  
 5170 Clusters in the Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates.  
 5171 *Mar. Drugs* 16, 67. doi:10.3390/md16020067.
- 5172 Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., et al.  
 5173 (2015). Plastic waste inputs from land into the ocean. *Science*. 347, 768–771.  
 5174 doi:10.1126/science.1260352.
- 5175 Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., et al. (2018). Structural  
 5176 insight into molecular mechanism of poly(ethylene terephthalate) degradation.  
 5177 *Nat. Commun.* 9, 382. doi:10.1038/s41467-018-02881-1.
- 5178 Karimi, E., Slaby, B. M., Soares, A. R., Blom, J., Hentschel, U., and Costa, R. (2018).  
 5179 Metagenomic binning reveals versatile nutrient cycling and distinct adaptive  
 5180 features in alphaproteobacterial symbionts of marine sponges. *FEMS*  
 5181 *Microbiol. Ecol.* 94. doi:10.1093/femsec/fiy074.
- 5182 Kawai, F., Kawabata, T., and Oda, M. (2019). Current knowledge on enzymatic PET  
 5183 degradation and its possible application to waste stream management and  
 5184 other fields. *Appl. Microbiol. Biotechnol.*, 1–16. doi:10.1007/s00253-019-09717-y.
- 5185 Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., et al. (2009).  
 5186 Isolation and Analysis of Bacteria with Antimicrobial Activities from the  
 5187 Marine Sponge *Haliclona simulans* Collected from Irish Waters. *Mar.*  
 5188 *Biotechnol.* 11, 384–396. doi:10.1007/s10126-008-9154-1.
- 5189 Kremling, A., Geiselmann, J., Ropers, D., and de Jong, H. (2015). Understanding  
 5190 carbon catabolite repression in *Escherichia coli* using quantitative models.  
 5191 *Trends Microbiol.* 23, 99–109. doi:10.1016/j.tim.2014.11.002.
- 5192 Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., et al.  
 5193 (2018). Total Synthesis of the Nonribosomal Peptide Surugamide B and  
 5194 Identification of a New Offloading Cyclase Family. *Angew. Chemie Int. Ed.* 57,  
 5195 9447–9451. doi:10.1002/anie.201805541.
- 5196 Labeda, D. P., Doroghazi, J. R., Ju, K.-S., and Metcalf, W. W. (2014). Taxonomic  
 5197 evaluation of *Streptomyces albus* and related species using multilocus  
 5198 sequence analysis and proposals to emend the description of *Streptomyces*

- 5199        albus and describe *Streptomyces pathocidini* sp. nov. *Int. J. Syst. Evol.*  
5200        *Microbiol.* 64, 894–900. doi:10.1099/ijs.0.058107-0.
- 5201        Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K.-S., et al.  
5202        (2017). Phylogenetic relationships in the family Streptomycetaceae using  
5203        multi-locus sequence analysis. *Antonie Van Leeuwenhoek* 110, 563–583.  
5204        doi:10.1007/s10482-016-0824-0.
- 5205        Lebreton, L., and Andrady, A. (2019). Future scenarios of global plastic waste  
5206        generation and disposal. *Palgrave Commun.* 5, 6. doi:10.1057/s41599-018-0212-  
5207        7.
- 5208        Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., et al. (2018). Protein Crystallography  
5209        and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate)  
5210        Hydrolase PETase from *Ideonella sakaiensis*. *ChemBioChem* 19, 1471–1475.  
5211        doi:10.1002/cbic.201800097.
- 5212        Magnus, N., Weise, T., and Piechulla, B. (2017). Carbon Catabolite Repression  
5213        Regulates the Production of the Unique Volatile Sodorifen of *Serratia*  
5214        *plymuthica* 4Rx13. *Front. Microbiol.* 8, 2522. doi:10.3389/fmicb.2017.02522.
- 5215        Matsuda, K., Kuranaga, T., Sano, A., Ninomiya, A., Takada, K., and Wakimoto, T.  
5216        (2019). The Revised Structure of the Cyclic Octapeptide Surugamide A. *Chem.*  
5217        *Pharm. Bull.* 67, 476–480. doi:10.1248/cpb.c19-00002.
- 5218        Mohr, K. I. (2016). “History of Antibiotics Research,” in (Springer, Cham), 237–272.  
5219        doi:10.1007/82\_2016\_499.
- 5220        Murphy, C. A., Cameron, J. A., Huang, S. J., and Vinopal, R. T. (1996). *Fusarium*  
5221        polycaprolactone depolymerase is cutinase. *Appl. Environ. Microbiol.* 62, 456–  
5222        60. doi:0099-2240/96/.
- 5223        Nawaz, A., Hasan, F., and Shah, A. A. (2015). Degradation of poly( $\epsilon$ -caprolactone)  
5224        (PCL) by a newly isolated *Brevundimonas* sp. strain MRL-AN1 from soil.  
5225        *FEMS Microbiol. Lett.* 362, 1–7. doi:10.1093/femsle/fnu004.
- 5226        Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., et al.  
5227        (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an  
5228        Unrelated Decapeptide, Surugamide F. *ChemBioChem* 17, 1709–1712.  
5229        doi:10.1002/cbic.201600350.
- 5230        Nishida, H., Suzuki, S., and Tokiwa, Y. (1998). Distribution of Poly( $\beta$ -propiolactone)  
5231        Aerobic Degrading Microorganisms in Different Environments. *J. Polym.*  
5232        *Environ.* 6, 43–58. doi:10.1023/A:1022878512743.

- 5233 Nishida, H., and Tokiwa, Y. (1993). Distribution of poly(?-hydroxybutyrate) and  
5234 poly(?-caprolactone)aerobic degrading microorganisms in different  
5235 environments. *J. Environ. Polym. Degrad.* 1, 227–233. doi:10.1007/BF01458031.
- 5236 Pan, R., Bai, X., Chen, J., Zhang, H., and Wang, H. (2019). Exploring Structural  
5237 Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A  
5238 Literature Review. *Front. Microbiol.* 10, 294. doi:10.3389/fmicb.2019.00294.
- 5239 Penn, K., and Jensen, P. R. (2012). Comparative genomics reveals evidence of  
5240 marine adaptation in *Salinispora* species. *BMC Genomics* 13, 86.  
5241 doi:10.1186/1471-2164-13-86.
- 5242 Romano, S., Jackson, S., Patry, S., and Dobson, A. (2018). Extending the “One Strain  
5243 Many Compounds” (OSMAC) Principle to Marine Microorganisms. *Mar.*  
5244 *Drugs* 16, 244. doi:10.3390/md16070244.
- 5245 Romero-Rodríguez, A., Rocha, D., Ruiz-Villafán, B., Guzmán-Trampe, S.,  
5246 Maldonado-Carmona, N., Vázquez-Hernández, M., et al. (2017). Carbon  
5247 catabolite regulation in *Streptomyces*: new insights and lessons learned. *World*  
5248 *J. Microbiol. Biotechnol.* 33, 0. doi:10.1007/s11274-017-2328-0.
- 5249 Sebastian, M., and Ammerman, J. W. (2009). The alkaline phosphatase PhoX is  
5250 more widely distributed in marine bacteria than the classical PhoA. *ISME J.* 3,  
5251 563–572. doi:10.1038/ismej.2009.10.
- 5252 Stülke, J., and Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr.*  
5253 *Opin. Microbiol.* 2, 195–201. doi:10.1016/S1369-5274(99)80034-4.
- 5254 Sun, W., Liu, C., Zhang, F., Zhao, M., and Li, Z. (2018). Comparative genomics  
5255 provides insights into the marine adaptation in sponge-derived *Kocuria flava*  
5256 S43. *Front. Microbiol.* 9, 1257. doi:10.3389/fmicb.2018.01257.
- 5257 Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., et al. (2013).  
5258 Surugamides A–E, Cyclic Octapeptides with Four <scp>d</scp> -Amino Acid  
5259 Residues, from a Marine *Streptomyces* sp.: LC–MS-Aided Inspection of Partial  
5260 Hydrolysates for the Distinction of <scp>d</scp> - and <scp>l</scp> -Amino  
5261 Acid Residues in the Sequence. *J. Org. Chem.* 78, 6746–6750.  
5262 doi:10.1021/jo400708u.
- 5263 Taylor, M. W., Radax, R., Steger, D., and Wagner, M. (2007). Sponge-associated  
5264 microorganisms: evolution, ecology, and biotechnological potential. *Microbiol.*  
5265 *Mol. Biol. Rev.* 71(2), 295–347. doi:10.1128/MMBR.00040-06.
- 5266 Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M. E., and Seipke, R. F.  
5267 (2019). A *trans*-Acting Cyclase Offloading Strategy for Nonribosomal Peptide  
5268 Synthetases. *ACS Chem. Biol.* 14, 845–849. doi:10.1021/acscchembio.9b00095.



- 5269 Tian, X., Zhang, Z., Yang, T., Chen, M., Li, J., Chen, F., et al. (2016). Comparative  
5270 Genomics Analysis of *Streptomyces* Species Reveals Their Adaptation to the  
5271 Marine Environment and Their Diversity at the Genomic Level. *Front.*  
5272 *Microbiol.* 7, 998. doi:10.3389/fmicb.2016.00998.
- 5273 Tierrafría, V. H., Ramos-Aboites, H. E., Gosset, G., and Barona-Gómez, F. (2011).  
5274 Disruption of the siderophore-binding desE receptor gene in *Streptomyces*  
5275 coelicolor A3(2) results in impaired growth in spite of multiple iron-  
5276 siderophore transport systems. *Microb. Biotechnol.* 4, 275–85.  
5277 doi:10.1111/j.1751-7915.2010.00240.x.
- 5278 Watve, M., Tickoo, R., Jog, M., and Bhole, B. (2001). How many antibiotics are  
5279 produced by the genus *Streptomyces*? *Arch. Microbiol.* 176, 386–390.  
5280 doi:10.1007/s002030100345.
- 5281 Wei, R., and Zimmermann, W. (2017). Biocatalysis as a green route for recycling the  
5282 recalcitrant plastic polyethylene terephthalate. *Microb. Biotechnol.* 10, 1302–  
5283 1307. doi:10.1111/1751-7915.12714.
- 5284 Xu, F., Nazari, B., Moon, K., Bushin, L. B., and Seyedsayamdost, M. R. (2017).  
5285 Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074  
5286 Using High-Throughput Elicitor Screens. *J. Am. Chem. Soc.* 139, 9203–9212.  
5287 doi:10.1021/jacs.7b02716.
- 5288 Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., et al.  
5289 (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate).  
5290 *Science*. 351, 1196–1199. doi:10.1126/science.aad6359.

# Appendix I

## Other publications and contributions

5291 1. Almeida, E. L., Margassery, L. M., Kennedy, J., & Dobson, A. (2018). Draft  
5292 Genome Sequence of the Antimycin-Producing Bacterium *Streptomyces* sp. Strain  
5293 SM8, Isolated from the Marine Sponge *Haliclona simulans*. *Genome*  
5294 *Announcements*, 6(4), e01535-17. doi:10.1128/genomeA.01535-17.

5295 **Abstract:** *Streptomyces* sp. strain SM8, isolated from *Haliclona simulans*, possesses  
5296 antifungal and antibacterial activities and inhibits the calcineurin pathway in yeast.  
5297 The draft genome sequence is 7,145,211 bp, containing 5,929 predicted coding  
5298 sequences. Several secondary metabolite biosynthetic gene clusters are present,  
5299 encoding known and novel metabolites, including antimycin.

5300 **Contribution:** Eduardo L. Almeida processed the 454 sequencing data obtained  
5301 using a GS FLX Titanium system (Roche), assembled and annotated the  
5302 *Streptomyces* sp. SM8 genome, performed quality assessments, data analysis, and  
5303 wrote the manuscript.

5304 2. Almeida, E. L., Margassery, L. M., O'Leary, N., & Dobson, A. (2018). Draft  
5305 Genome Sequence of *Pseudomonas putida* CA-3, a Bacterium Capable of Styrene  
5306 Degradation and Medium-Chain-Length Polyhydroxyalkanoate Synthesis.  
5307 *Genome Announcements*, 6(4), e01534-17. doi:10.1128/genomeA.01534-17.

5308 **Abstract:** *Pseudomonas putida* strain CA-3 is an industrial bioreactor isolate  
5309 capable of synthesizing biodegradable polyhydroxyalkanoate polymers via the  
5310 metabolism of styrene and other unrelated carbon sources. The pathways involved  
5311 are subject to regulation by global cellular processes. The draft genome sequence is  
5312 6,177,154 bp long and contains 5,608 predicted coding sequences.

5313 **Contribution:** Eduardo L. Almeida processed the Illumina's MiSeq paired-end  
5314 sequencing data, assembled and annotated the *Pseudomonas putida* strain CA-3  
5315 genome, performed quality assessments, data analysis, and wrote the manuscript.

5316 **3.** Gil-Pulido, B., Tarpey, E., Almeida, E. L., Finnegan, W., Zhan, X., Dobson, A., &  
 5317 O'Leary, N. (2018). Evaluation of dairy processing wastewater biotreatment in an  
 5318 IASBR system: Aeration rate impacts on performance and microbial ecology.  
 5319 *Biotechnology Reports (Amsterdam, Netherlands)*, 19, e00263.  
 5320 doi:10.1016/j.btre.2018.e00263.

5321 **Authors list:** Beatriz Gil-Pulido, Emma Tarpey, Eduardo L. Almeida, William  
 5322 Finnegan, Xinmin Zhan, Alan D.W. Dobson, and Niall O'Leary.

5323 **Abstract:** Dairy processing generates large volumes of wastewater that require  
 5324 extensive nutrient remediation prior to discharge. Significant commercial  
 5325 opportunities exist therefore for cost-effective biotechnologies capable of achieving  
 5326 this requirement. In this study the authors evaluated the use of intermittently  
 5327 aerated sequencing batch reactors, (IASBRs), as a single-tank biotreatment system  
 5328 for co-removal of COD, nitrogen and phosphorus from synthetic dairy processing  
 5329 wastewater. Variation of the IASBR aeration rates, (0.8, 0.6 and 0.4 L/min), had  
 5330 significant impacts on the respective nutrient removal efficiencies and underlying  
 5331 microbial diversity profiles. Aeration at 0.6 L/min was most effective and resulted  
 5332 in >90% co-removal of orthophosphate and ammonium. 16S rRNA based  
 5333 pyrosequencing of biomass DNA samples revealed the family *Comamonadaceae*  
 5334 was notably enriched (>80% relative abundance) under these conditions. *In silico*  
 5335 predictive metabolic modelling also identified *Comamonadaceae* as the major  
 5336 contributor of several known genes for nitrogen and phosphorus assimilation (*nirK*,  
 5337 *nosZ*, *norB*, *ppK*, *ppX* and *phbC*).

5338 **Contribution:** Eduardo L. Almeida performed the *in silico* metabolic modelling  
 5339 analysis reported in this work, employing 16S amplicon analysis tools such as  
 5340 QIIME and PICRUST, together with statistical analysis such as Principal  
 5341 Component Analysis.

5342 4. Jackson, S. A., Crossman, L., Almeida, E. L., Margassery, L. M., Kennedy, J., & Dobson, A.  
 5343 (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the  
 5344 Genomes of Marine Sponge Derived *Streptomyces* spp. Isolates. *Marine Drugs*, 16(2), 67.  
 5345 doi:10.3390/md16020067.

5346 **Abstract:** The genus *Streptomyces* produces secondary metabolic compounds that are rich in  
 5347 biological activity. Many of these compounds are genetically encoded by large secondary  
 5348 metabolism biosynthetic gene clusters (smBGCs) such as polyketide synthases (PKS) and non-  
 5349 ribosomal peptide synthetases (NRPS) which are modular and can be highly repetitive. Due to  
 5350 the repeats, these gene clusters can be difficult to resolve using short read next generation  
 5351 datasets and are often quite poorly predicted using standard approaches. We have sequenced  
 5352 the genomes of 13 *Streptomyces* spp. strains isolated from shallow water and deep-sea  
 5353 sponges that display antimicrobial activities against a number of clinically relevant bacterial  
 5354 and yeast species. Draft genomes have been assembled and smBGCs have been identified  
 5355 using the antiSMASH (antibiotics and Secondary Metabolite Analysis Shell) web platform. We  
 5356 have compared the smBGCs amongst strains in the search for novel sequences conferring the  
 5357 potential to produce novel bioactive secondary metabolites. The strains in this study recruit to  
 5358 four distinct clades within the genus *Streptomyces*. The marine strains host abundant smBGCs  
 5359 which encode polyketides, NRPS, siderophores, bacteriocins and lantipeptides. The deep-sea  
 5360 strains appear to be enriched with gene clusters encoding NRPS. Marine adaptations are  
 5361 evident in the sponge-derived strains which are enriched for genes involved in the  
 5362 biosynthesis and transport of compatible solutes and for heat-shock proteins. *Streptomyces*  
 5363 spp. from marine environments are a promising source of novel bioactive secondary  
 5364 metabolites as the abundance and diversity of smBGCs show high degrees of novelty. Sponge  
 5365 derived *Streptomyces* spp. isolates appear to display genomic adaptations to marine living  
 5366 when compared to terrestrial strains.

5367 **Contribution:** Eduardo L. Almeida contributed with the bioinformatics-based analyses present  
 5368 in this paper, particularly those related to the analysis of secondary metabolism protein  
 5369 domains of interest present in the marine *Streptomyces* spp. analysed in this work, employing  
 5370 the antiSMASH program together with custom Python and R scripts for data processing.